OLUME 103 PART B NUMBER 8 MARCH 1956

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE . LONDON W.C.2

Price Seven Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1955-56

President

SIR GEORGE H. NELSON, Bart.

Past-Presidents

SIR JAMES SWINBURNE, Bart., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc. (Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc. (Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
J. ECCLES, C.B.E., B.Sc.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
C. M. COCK.
A. R. COOPER, M.Eng.
A. T. CRAWFORD, B.Sc.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
D. MCDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.
H. H. MULLENS, B.Sc.
W. F. PARKER.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. L. WATES, J.P.
G. O. WATSON.
D. B. WELBOURN, M.A.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairman and Past-Chairmen of Sections

Measurement and Control:
W. BAMFORD, B.Sc.
*M. WHITEHEAD.

Radio and Telecommunication:
H. STANESBY.
*C. W. OATLEY, O.B.E., M.A., M.Sc.

Supply:
L. DRUCQUER.
*J. D. PEATTIE, B.Sc.

Utilization:
D. B. HOGG, M.B.E., T.D.
*J. I. BERNARD, B.Sc.Tech.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
F. R. C. ROBERTS.
*J. M. MITCHELL, B.Sc., Ph.D.

Mersey and North Wales Centre:
PROFESSOR J. M. MEEK, D.Eng.
*P. R. DUNN, B.Sc.

North Midland Centre:
F. BARRELL.
*G. CATON.

North-Eastern Centre:
A. H. KENYON.
*G. W. B. MITCHELL, B.A.

Southern Centre:
L. H. FULLER, B.Sc.(Eng.).
*E. A. LOGAN, M.Sc.

North-Western Centre:
G. V. SADLER.
*PROFESSOR E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.

Northern Ireland Centre:
MAJOR E. N. CUNLIFFE, B.Sc.Tech.
*MAJOR P. L. BARKER, B.Sc.

Western Centre:
T. G. DASH, J.P.
*A. N. IRENS.

Scottish Centre:
*E. WILKINSON, Ph.D., B.Eng.
*J. S. HASTIE, B.Sc.(Eng.).

South Midland Centre:
H. S. DAVIDSON, T.D.
*A. R. BLANDFORD.

* Past-Chairman.

RADIO AND TELECOMMUNICATION SECTION COMMITTEE 1955-56

Chairman

H. STANESBY

Vice-Chairmen

R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
J. S. MCPETRIE, Ph.D., D.Sc.

Past-Chairmen

C. W. OATLEY, O.B.E., M.A., M.Sc.
J. A. SMALE, C.B.E., A.F.C., B.Sc.

Ordinary Members of Committee

PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
F. S. BARTON, C.B.E., M.A., B.Sc.
A. J. BIGGS, Ph.D., B.Sc.
G. G. MACFARLANE, Dr.Ing., B.Sc.
B. N. MACLARTY, O.B.E.
H. PAGE, M.Sc.
N. C. ROBERTSON, C.M.G., M.B.E.
W. ROSS, M.A.
L. RUSHFORTH, M.B.E., B.Sc.
T. B. D. TERRONI, B.Sc.
A. M. THORNTON, B.Sc.
F. WILLIAMS, B.Sc.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.) (representing the Council).
E. H. COOKE-YARBOROUGH (Co-opted Member).
BRIG. E. J. H. MOPPETT (representing the Cambridge Radio and Telecommunication Group).
J. MOIR (representing the South Midland Radio and Telecommunication Group).
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.) (representing the North-Eastern Radio and Measurements Group).

The following nominees of Government Departments:
Admiralty: CAPTAIN G. C. F. WHITAKER, R.N.
Air Ministry: AIR COMMODORE G. H. RANDLE, R.A.F., B.A.
Department of Scientific and Industrial Research: B. G. PRESSEY, M.Sc.(Eng.), Ph.D.
Ministry of Supply: BRIG. J. D. HAIGH, O.B.E., M.A.
Post Office: CAPTAIN C. F. BOOTH, O.B.E.
War Office: COL. E. I. E. MOZLEY, M.A.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

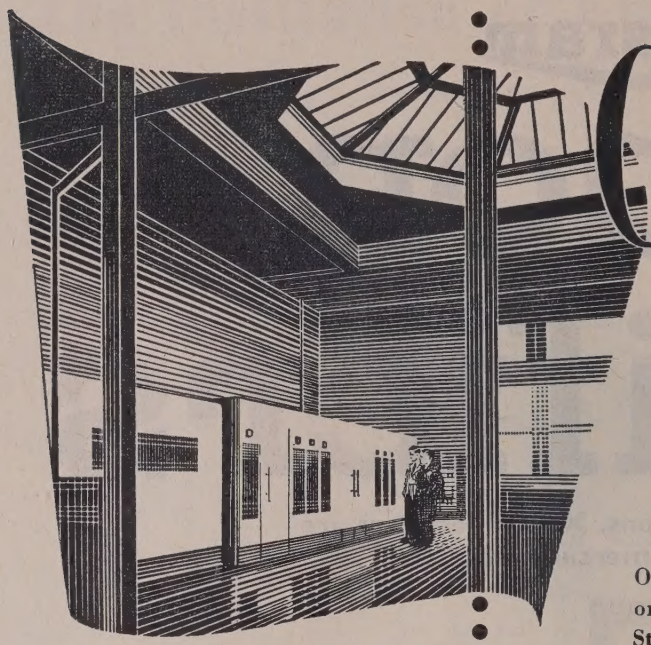
F. JERVIS SMITH, M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.



Communications by Standard

One of the largest telecommunication engineering organisations in the British Commonwealth **Standard Telephones and Cables Limited** is engaged in research, and the development, manufacture and installation of all types of communication and control systems. Concerned with every aspect of telecommunications engineering, the Company is in an unrivalled position to undertake, within its own organisation, the co-ordinated planning of complete communication projects involving inter-dependent systems of various types.

'Standard' products include:—

Telecommunication Line Transmission Equipment
Radio Broadcasting Equipment
Radio Communication Equipment
Air Radio Navigational Aids
Supervisory and Remote Control Systems
Railway Communication Apparatus
Railway Control Equipments
Telephone Cable
Sound-Reinforcement Systems
Public and Private Telephone Systems

☆

'Standard' telecommunication systems serve society on land, at sea and in the air. In the field of radio broadcasting 'Standard' transmitters are in service with leading broadcasting administrations throughout the world. Illustrated is a 'Standard' CM.10 200kW broadcasting transmitter at the British Broadcasting Corporation Station at Washford Cross in Somerset.

Standard Telephones and Cables Limited
TELECOMMUNICATION ENGINEERS

CONNAUGHT HOUSE, ALDWYCH, LONDON, W.C.2, ENGLAND
CAIRO · COLOMBO · DUBLIN · JOHANNESBURG · KARACHI · NEW DELHI · SALISBURY

G.E.C.

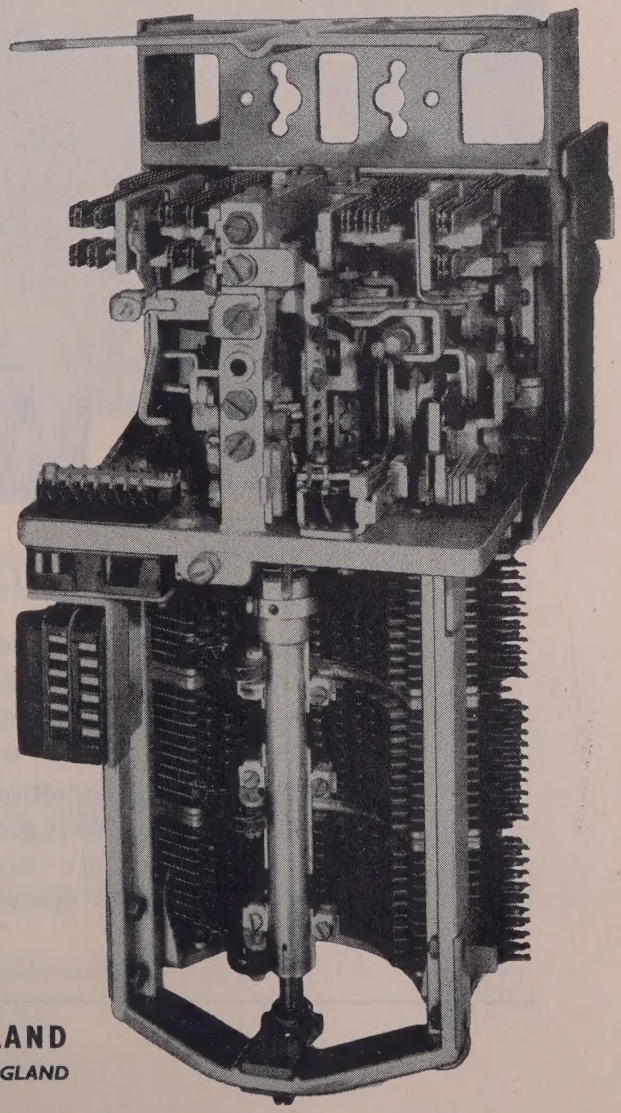
SE50 Selector

Local technician at work on the new exchange at Kuching—the first exchange to use SE50 Selectors in S.E. Asia.

. . . now cutting maintenance costs in South-East Asia!

THE NEW automatic exchange at Kuching, Sarawak, in the Island of Borneo, is the first main exchange using SE50 Selectors to be commissioned in South-East Asia. The initial capacity of the exchange is 800 subscribers but it is designed to have a capacity of 2,000 subscribers.

SE50 Selectors have been tested to five times the British Post Office requirements for two-motion Selectors, and their remarkable life, extraordinary reliability and ease of maintenance render them particularly suitable for an installation of this kind.



The SE50 sets new standards in automatic exchange equipment. The mechanism is robust and simple; all adjustments are independent and can be made without disturbing those previously carried out. These characteristics mean that the SE50 can be maintained by less-skilled staff—providing yet another saving in maintenance costs.

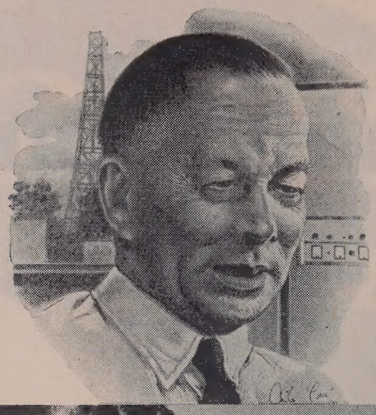
★ Each unit can be removed from the frame for adjustment and replaced without affecting its own or other settings.

★ Drift adjustments are eliminated and all adjustments are independent.

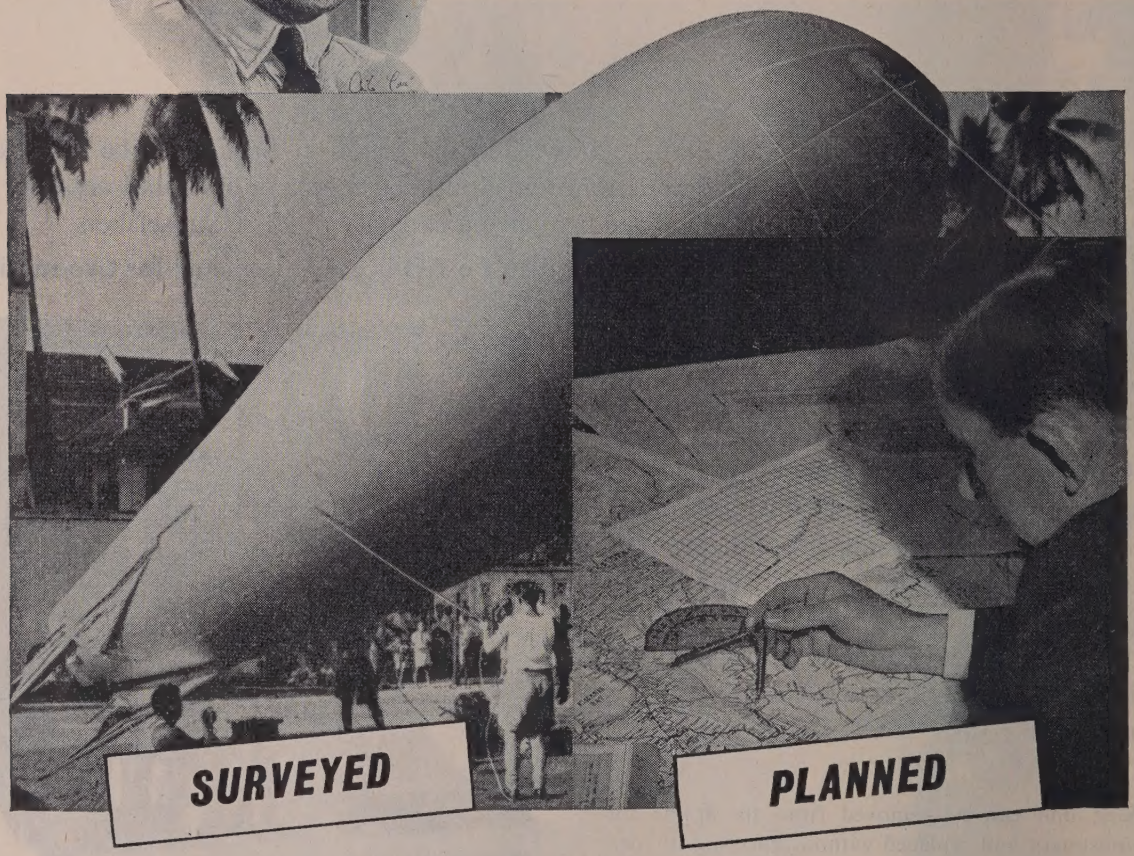
★ Auxiliary contact spring/sets are identical with those used in G.E.C. P.10 type relays (British Post Office 600 type). No special adjustment technique is required.

THE GENERAL ELECTRIC CO LTD OF ENGLAND

TELEPHONE, RADIO AND TELEVISION WORKS, COVENTRY, ENGLAND



MARCONI



V. H. F. MULTI-CHANNEL SYSTEMS

VHF radio-telephony was welcomed in its early days as an economic means of providing communication over terrain where the cost and difficulty of erecting line or cable routes was prohibitive. Today Marconi's have developed multi-channel systems which, by employing frequency modulation to widen the bands, provide up to 48 telephone channels and are recognised as being preferable to line systems in many instances on grounds of performance as well as installation cost.

Our first customer was a Lighthouse

Twenty five years ago we needed stand-by generating plant for Chance Lighthouses. In view of the quite exceptional reliability demanded, we decided we had better design and build the plant ourselves. That was how we came into Austinlite.

But many other things require this same kind of reliability: telecommunications, for example. And so you find Austinlite plant supplying an unbroken flow of power to telephone systems, radio and radar beacons, airfield control and landing aids in all parts of the World.

This reliability for which Austinlite has become famous is the result of two things. An inbuilt refinement of design, and the exact adaptation of the plant to the job. That is why, beside standard Austinlite plant, much Austinlite equipment is tailor made. Give Austinlite engineers a clear idea of the difficulties to be overcome, and provided the job is practicable at all, they will design plant to overcome them. They have done so again and again.

Austinlite **AUTOMATIC GENERATING PLANT**

Tailor-made by STONE-CHANCE LTD.

(The makers of Sumo Pumps & Stone-Chance Lighthouses) 28 St. James's Square, London, S.W.1 · Tel.: Trafalgar 1954

FERRANTI ELECTRONIC DIGITAL COMPUTERS

FERRANTI MARK 1* COMPUTER

FERRANTI PEGASUS COMPUTER

Ferranti Ltd. are making computers suitable for a wide variety of applications in the fields of mathematics, engineering, science, commerce, accountancy and general statistics. We will investigate the problems of using computers in any of these fields, and advise on the specification of suitable machines.

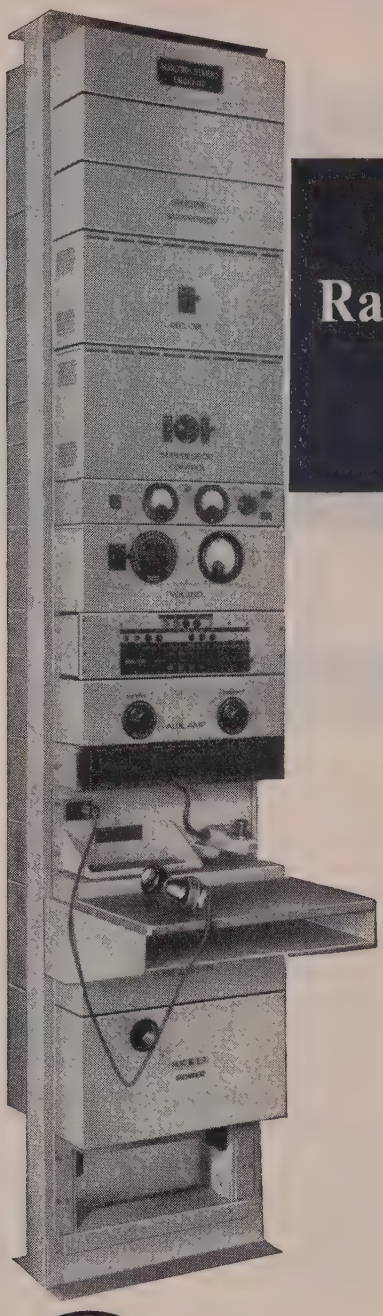
Write for full information to Ferranti Ltd., Computer Dept., Moston, Manchester 10, or the London Computer Centre, 21 Portland Place, London, W.1.

FERRANTI MERCURY COMPUTER

FERRANTI LTD · MOSTON · MANCHESTER 10

London Computer Centre: 21 PORTLAND PLACE, W.1.

FERRANTI



MARCONI-SIEMENS
Radio Telephone Terminal
TYPE B
(HW 21)

The type B (HW 21) terminal provides a satisfactory junction of HF radio with line or cable telephone and telegraph circuits. Its primary function is to eliminate the unstable conditions due to the inherently high gain in the radio link by ensuring that the radio circuit is operative in one direction only at any one instant. It also provides facilities for controlling the signal levels to the line or to the radio transmitters for discriminating against line and radio noises, and for simple privacy working. Its features include semi-automatic operation, two or four-wire line connection, electronic VF switching, radio calling facilities, and centralised test and monitoring facilities. It is self-contained for AC mains supply.

Marconi

THE LINK BETWEEN RADIO AND LINE COMMUNICATIONS

SIEMENS BROTHERS

Full details of this and other Marconi-Siemens equipment, which provides completely integrated radio and line telegraph and telephone systems may be obtained from either—

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX
OR SIEMENS BROTHERS & CO., LIMITED, WOOLWICH, LONDON, S.E.18

MS 1

New!

E.H.T. CAPACITOR

Type 500 H.T. oil-filled capacitor in non-metallic case

PLASTIC CASE ELIMINATES CORONA EFFECTS

HERMETICALLY SEALED

SINGLE HOLE FIXING

Terminations		$\frac{7}{16}$" x 4 B.A.		$\frac{7}{16}$" x 4 B.A.		$\frac{1}{2}$" x 2 B.A.	
Tube Size		Dia.	Len.	Dia.	Len.	Dia.	Len.
D.C. Rating 70° C. Wkg.	Test	$\frac{13}{16}$"	$2\frac{3}{8}$"	$1\frac{1}{8}$"	$2\frac{3}{8}$"	$1\frac{1}{8}$"	$3\frac{1}{8}$"
2.0 kV.	4.0 kV.	0.05 μF.		0.1 μF.		0.15 μF.	
4.0 kV.	8.0 kV.	0.01 μF.		0.02 μF.		0.03 μF.	
5.0 kV.	10.0 kV.	0.0075 μF.		0.015 μF.		0.02 μF.	
8.0 kV.	16.0 kV.	0.002 μF.		0.005 μF.		0.0075 μF.	
10.0 kV.	20.0 kV.	0.0015 μF.		0.003 μF.		0.005 μF.	
12.0 kV.	24.0 kV.	0.001 μF.		0.002 μF.		0.003 μF.	
15.0 kV.	30.0 kV.	0.0007 μF.		0.0015 μF.		0.002 μF.	
18.0 kV.	36.0 kV.	0.0005 μF.		0.001 μF.		0.0015 μF.	
20.0 kV.	40.0 kV.	0.0003 μF.		0.0007 μF.		0.001 μF.	
25.0 kV.	50.0 kV.	0.0002 μF.		0.0005 μF.		0.00075 μF.	
30.0 kV.	60.0 kV.	—		—		0.0005 μF.	

The capacitance value is a nominal maximum for a given tube size and voltage rating, capacitance tolerance ± 20%.

DUBILIER

DUBILIER CONDENSER CO. (1925) LTD., DUCON WORKS, VICTORIA RD., NORTH ACTON, W.3.
Telephone: ACOrn 2241 Telegrams: Hivoltcon, Wesphone, London.

DN150A

High Capacitance Plate Ceramicons*

This trade mark, which is associated particularly with ERIE CERAMICONS, is recognised throughout the world as a symbol of quality and reliability.

The capacitor bearing the trade mark is the type CP3/100, the first in a new range of HIGH CAPACITANCE PLATE CERAMICONS.

This modest coin is intended to indicate comparative physical size and to convey the fact that the price is equally modest.

ERIE CERAMICONS are renowned for their long life, good ageing factor, high insulation resistance and full climatic protection, all of which are ensured and maintained by exacting engineering control in every stage of manufacture from raw material to finished product.

These good qualities, coupled with the brief specification given below, will commend the ERIE type CP3/100 PLATE CERAMICON to the equipment designer, who is seeking reliability and high capacitance in the smallest possible physical size, at a really economic price :—

ERIE TYPE CP3/100 PLATE CERAMICON

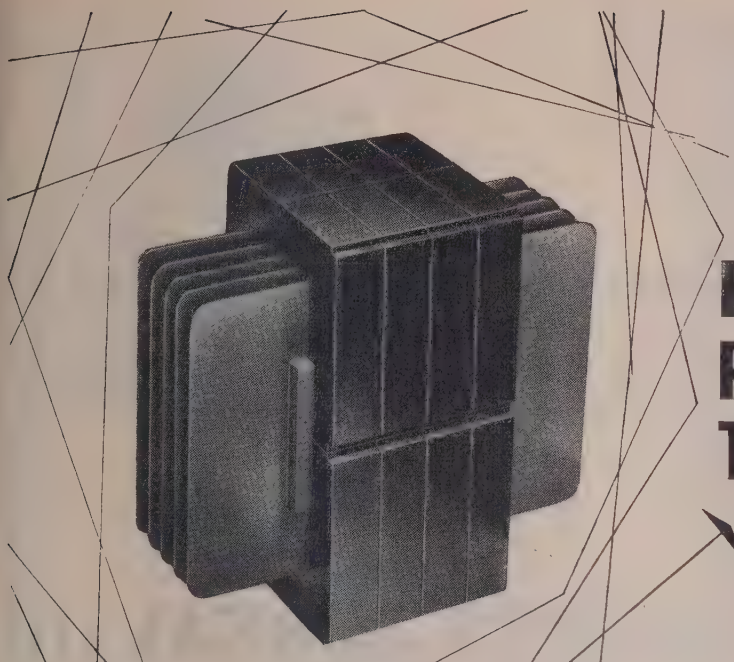
0.03 mfd. G.M.V. or + 80%, — 20%
500 Volts D.C. Working. 350 Volts A.C. Working.
Dimensions: 0.850″ × 0.850″ × 0.200″

The ERIE type CP3/100 PLATE CERAMICON, like other ERIE CERAMICONS, is made from start to finish, in our factory at Great Yarmouth, and bulk supplies are immediately available.

ERIE* *Resistor Ltd*

* Registered Trade Marks

Carlisle Road, The Hyde, London, N.W.9., England. Telephone: COLindale 8011. *Factories: London and Great Yarmouth, England; Trenton, Ont., Canada; Erie, Pa., and Holly Springs, Miss., U.S.A.*



H.F. POWER TRANSFORMERS

RATING . . . . . UP TO 2 kW

FREQUENCY RANGE . . . . 2 Kc/s to 2 Mc/s

H.F. power transformers of outstanding efficiency are the latest addition to the Mullard range of high quality components designed around Ferroxcube magnetic cores.

Utilising the unique characteristics of Ferroxcube to the full, Mullard H.F. transformers are smaller, lighter, and less costly than transformers using alternative core materials. These advantages are particularly marked in transformers required to handle powers of up to 2kW, between the frequency range 2kc/s to 2Mc/s.

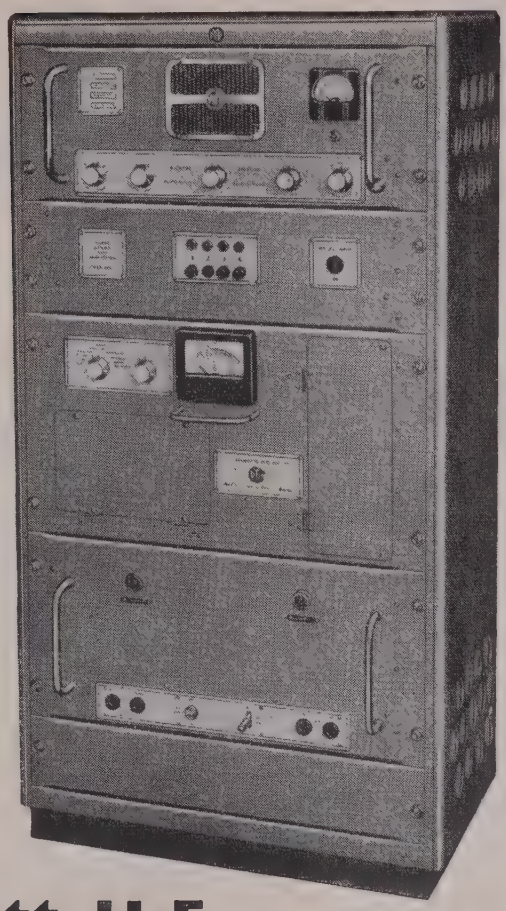
Mullard transformers are already finding wide use in applications as diverse as ultrasonic H.F. power generators and aircraft power packs operating from an aircraft's normal A.C. supply. In the latter application, the low leakage field of Ferroxcube can eliminate the need for external screening, thereby reducing the size and weight of the transformer even further.

As with all Mullard high quality components, these H.F. power transformers are designed and built to engineers' individual specifications. Write now for details of the complete range of components available under this service.

Mullard

'Ticonal' permanent magnets
Magnadur ceramic magnets
Ferroxcube magnetic cores

MULLARD LTD · COMPONENTS DIVISION · CENTURY HOUSE · SHAFTESBURY AVENUE · WC2
MC 1



60 Watt H.F. FIXED STATION

This completely new Pye equipment has been specifically designed for point-to-point communication and will fulfil equally well a ground-to-air role in air traffic control systems.

Push button control brings any one of four preselected channels into immediate operation; this facility is also available when the equipment is installed for remote unattended operation. The 60 watt Fixed Station Transmitter offers R/T, C/W, or M.C.W. operation with 'break-in' facilities on telegraphy.

The equipment is suitable for unattended operation in the tropics.

Telecommunications

CAMBRIDGE ENGLAND

Pye (New Zealand) Ltd. Auckland C.I., New Zealand	Pye Canada Ltd. Ajax, Canada	Pye-Electronic Pty., Ltd. Melbourne, Australia	Pye (Ireland), Ltd. Dublin, Eire
Pye Radio & Television (Pty.) Ltd. Johannesburg South Africa	Pye Limited Mexico City	Pye Limited Tucuman 829 Buenos Aires	Pye Corporation of America 270 Park Avenue New York

PYE LIMITED • • CAMBRIDGE • • ENGLAND

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

TEN-YEAR INDEX

1942—1951

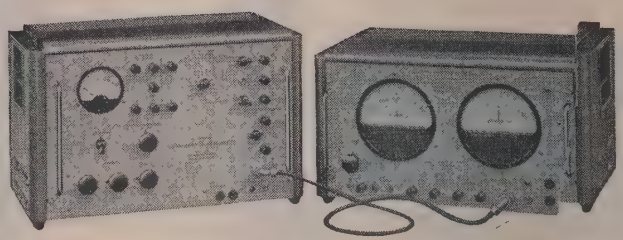
A TEN-YEAR INDEX to the *Journal of The Institution of Electrical Engineers* for the years 1942–48 and the *Proceedings* 1949–51 (vols. 89–98) can be obtained on application to the Secretary.

The published price is £1 5s. 0d. (post free), but any member of The Institution may have a copy at the reduced price of £1 (post free).

Danger—Top Secret!

Alas, as the day for firing draws near, Dr. Frittertime discovers that his missile is unstable. Months of patient endeavour have been wasted. If only Dr. Frittertime had used a Solartron Transfer Function Analyser, all his servo control problems would have been solved—and extremely swiftly too.

Write now for details of the many applications of this invaluable Solartron product.



- Tests A.C., D.C. or carrier servos
- Plots high accuracy Nyquist diagrams
- Covers 0·1 c/s to 1,000 c/s
- Independent of harmonics, noise & spurious frequencies
- Answers G.W., fire-control, simulator, computor, vibration, magamp and all servo problems

THE SOLARTRON ELECTRONIC GROUP LTD

RELIABILITY—UNDER OUR 12 MONTHS' GUARANTEE, COSTS HAVE NEVER EXCEEDED 0·2% OF SALES

Thames Ditton, Surrey
Telephone: EMBerbrook 5522
Cables: Solartron, Thames Ditton

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 103. PART B. NO. 8. MARCH 1956

DISCUSSION ON "A TRANSISTOR DIGITAL FAST MULTIPLIER WITH MAGNETOSTRICTIVE STORAGE"*

BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 9TH NOVEMBER, 1955

Mr. E. H. Cooke-Yarborough: There seems to be a tendency to dislike point-contact transistors and to think it unfashionable to use them, as one of the authors has said. I think this attitude is to be deplored and I am not sure why it exists. Unreliability is often quoted as a reason for not using these transistors. However, the figures quoted in the paper, which refer to transistors made several years ago, indicate that unreliability need no longer be a serious objection, if circuits are properly designed. I hesitate to suggest it, but I wonder whether some of the antipathy to point transistors may not arise from sheer laziness. Junction-transistor circuits bear quite a close family resemblance to valve circuits with which we are familiar, whereas point-transistor circuits are like nothing else, and have to be learnt fresh. Whatever the reason, I think we should make an effort to overcome this prejudice, because there is no doubt that very efficient simple circuits can be made with point transistors, as the authors have shown.

Perhaps I should comment on the authors' use of a base inductor. It has been said that this makes the transistor easier to turn off, and, given the right circumstances, that is so. But if we consider the circuit of Fig. 2, in which a defined base current is used, we know that this base current should not exceed about one-third of the minimum collector load current if the transistor is to be stable in the "on" condition. This is true also in the case of the circuit with the base inductor (Fig. 4).

At the beginning of a pulse the base current in this circuit is quite small, being I_{co} plus the current in the damping resistor. This current rises linearly during the pulse, but it must not exceed one-third of the minimum collector current during the time that the pulse must exist, or the transistor might turn-off prematurely. So if at the end we apply a turn-off pulse, this pulse has just as much work to do as it had in the circuit with the defined base current.

The circuit having the base inductor is advantageous when some tolerance can be allowed in the length of the pulse it delivers. We can now fix a time t_1 before which the circuit must not turn itself off, choose a value for the inductance such that the current does not exceed one-third of the collector current before t_1, and apply the turn-off pulse at a somewhat later time t_2. The base current continues to rise after t_1, bringing the transistor much nearer to its turn-off condition, and some transistors will turn-off spontaneously. Quite a light turn-off pulse applied at t_2 will now suffice to turn-off those transistors which are still conducting.

It is instructive to consider the problems which would arise if one tried to use junction transistors in a device of the sort described. Junction-transistor bi-stable circuits using pairs of ordinary hearing-aid-type transistors have been made to count at a rate of about 100 kc/s, and shifting registers have been made to operate at about this rate, but they have little in hand. The collector voltage has to be caught by diodes to prevent bottoming and thus avoid hole storage in the transistor. Even so, the pulse edges are slow, and the gating waveforms are therefore imprecise. There is little spare current available for operating gate circuits, because each transistor is doing all it can to turn the other off or on. An emitter follower may therefore be needed to feed external circuits, so three junction transistors may be needed to do a job that is done well by a single point-contact transistor.

Faster junction transistors are becoming available, but the voltage that can be applied to their collectors is restricted, because of the danger of punch-through between collector and emitter. These high-speed junction transistors tend, therefore, to be limited to 10 or even 4½ volts at the collector. Reduction of the collector voltage swing much below that referred to in the paper would lead to tolerance difficulties in the diode gating circuits, owing to the finite forward voltage drop which exists in available diodes.

It would be particularly hard to use a junction transistor in the sending circuit, Fig. 24, which must deliver a large voltage swing and a well-defined current pulse. Unfortunately, the collector voltage waveform has not been shown in the diagram, but to give a sharp cut-off of collector current the collector voltage probably swings down to at least −30 volts. Fast current pulse edges combined with swings of 30 volts or more are difficult to achieve with junction transistors, whether of the high- or low-speed variety. There is a recent American paper† which evaluates the best switching performance theoretically obtainable from alloyed junction transistors if an optimum design of transistor could be realized. The generation of a 30-volt, 0·1 microsec pulse edge is considered. It is shown that this could not quite be achieved with an ideal *p–n–p* transistor and could only just be achieved with an ideal *n–p–n* transistor. On the other hand,

* CHAPLIN, G. B. B., HAYES, R. E., and OWENS, A. R.: Paper No. 1858 R, July, 1955 (see 102 B, p. 412).

† EBERS, J. J., and MILLER, S. L.: "Design of Alloyed Junction Germanium Transistors for High Speed Switching," *Bell System Technical Journal*, 1955, 34, p. 780.

point-contact transistors available in this country to-day provide a performance almost as good as that quoted. The performance of junction transistors is likely always to fall short of the ideal, so there is a lot to be said for using point-contact transistors in the right circuits both now and in the future.

As there seems to be some disagreement about how transistor circuit diagrams should be presented, I should like to draw attention to the diagrams of this paper. The authors have followed the practice established for valve circuits and have put positive parts of the circuit in the upper parts of the diagram and negative parts below. They have also differentiated between point-contact and junction transistors by use of different symbols.

A diagram such as Fig. 26, for example, demonstrates, I think, that transistor circuits can gain much in clarity by adherence to these conventions. Is it too much to hope that some such conventions will be widely followed?

Mr. T. R. Scott: The authors must be congratulated on their very rational approach to the use of present-day point-contact transistors. By judicious circuit techniques they have overcome difficulties associated with the magnitude and variability of i_{co}, the magnitude of the turn-off time, and the reduction in life due to overloading. In due course the fruits of work now proceeding in various laboratories in connection with improved point-contact devices will be enjoyed commercially and there will be less need to adjust the circuit techniques to cover up the deficiencies of such transistors. It is, however, relatively clear that the user will always have to compromise between life and overload if he desires to apply overload. Of course at present we are not quite clear where overload commences, i.e. the load at which deterioration sets in. The extant data are largely empirical.

There is a tendency at present to write down the point-contact device as something only of historical interest, but I cannot see any simple alternative to it (diode as well as transistor) if we are to have really high-speed switching in the near future.

It is not very clear whether the authors settled their circuit technique, in terms of 5 adders, purely on the question of the cost ratio of transistor to diode. If this is so it would appear that their ratio is not in accordance with present-day costs. A ratio of 3 : 1 would be nearer the mark, and this would justify the use of 8 adders with very much simplified gates. However, as the authors point out, the complexity involved by their choice is a real test of the transistor and, as such, has served a useful purpose.

On the question of gates, Fig. 7 has a left-hand and a right-hand gate. The gate on the left hand seems to be preferable, if we can get rid of the stray capacitances due to wiring. I should be interested to know whether the authors have paid any attention to this question and what results they have obtained.

Mr. B. W. Pollard: Ever since the transistor was announced we have had a steady stream of papers concerning the design of transistor circuits. It would appear that most of these papers have been concerned with the setting-up of mathematical models for specific circuits.

There seems to have been very little generalization in these papers, but in the present paper the essentially simple nature of the generalized circuit design procedure has been well demonstrated, as has the logical design procedure, which has resulted in a rather complex logical device. I believe that there is no real difficulty in designing transistor circuits, providing that the fundamental parameters are well understood. Indeed I have the impression that many analyses of transistor circuits have ignored the true analogies with ordinary valve circuits.

Probably one of the main simplifications that the authors have achieved has been the bringing together of valve circuits and transistor circuits. The valves are used to produce standardized continuous waveforms, and these valve circuits, because of their nature, are easy to maintain and check. The transistors are used in the logical circuits where checking and maintenance is more difficult.

The authors have circumvented the difficulties of buildin complete computer by using another computer to prod answers which could be checked against those obtained by multiplier. However, a very large number of tests need to performed in order to test the system thoroughly, and I sho like to know whether the authors, in deciding upon this particu form of testing, considered any alternative procedures, or whet they felt this was the best way of testing such equipment.

Whilst the paper has been entirely concerned with po contact transistors, I feel that ultimately the junction transis will be of greater importance, since once its frequency ra has been extended, inherently greater reliability and rep ducibility should outweigh any other advantage that a poi contact transistor might have.

Very rapid rise-times may be achieved from point-cont transistors, but it is not a necessary condition of a digital circ that the rapid rise time should be achieved directly. One c achieve the requirements by making use of switching wavefor as the authors have done, and in fact the basic operating freque may well be increased by a factor of ten.

I am very pleased to see that the magnetostrictive delay li used have been operated from transistors, since I regard th delay lines as a very vital part of any new computer equipmer

Dr. R. L. Grimsdale: A relatively high pulse repetition f quency of 125 kc/s has been chosen. This is the same as that the transistor computer at Manchester University, which been built in parallel with this multiplier but as a distinct proje The choice of this high pulse repetition frequency is v important, because it has shown that the transistors can at le approach the operating speed used in present computers.

The authors have used circuits which are turned on at emitter and turned off or inhibited at the base. The "turn o is a particularly difficult feature owing to hole storage, and th is apparently some ageing process which makes it rather m difficult to turn off transistors after long periods of service.

From the authors' remarks, the use of base coils appe partly to overcome this difficulty. However, this technique m be used with care and indeed is not possible in those ca where we do not know how long the transistor is going remain on.

Regarding the question of junction versus point-cont transistors, the junction transistor appears more reliable than average point-contact transistor, but in very-high-speed circu the point-contact transistor is better at the moment. But I ag with previous speakers that we could operate junction transist at 50–100 kc/s with reasonably fast transitions, particularly if transistor connected as an earthed emitter amplifier is driv both on and off, with current from a negative source to drive on and current from a positive source to drive it off.

What is the overall delay in the reading amplifiers for magnetostriction lines?

Dr. J. H. Stephen: The authors state that they have hal digit margin of safety when strobing the output from a sto Do the characteristics of the line vary, or does the 120 k oscillator drift sufficiently, to necessitate readjustment of t timing, which is presumably done by altering the position of head on the line? It might be of interest to hear how lo (physically) these stores are; i.e. the distance d in Fig. 22(a).

The authors are using base inductances to time the on-state a transistor (Figs. 4, 10 and 24). We have found at A.E.R that the pulse length produced by this type of circuit is sensit to the characteristics of the transistor—particularly the curr gain and the collector current. The use of this type of circ (Fig. 24) for timing the sending waveform appears to be unsa

factory, since changes in the sending transistor will alter the shape of the received pulse and may upset the strobing. Have the authors any comments to make on this effect?

Both positive and negative 50-volt supplies are used for providing defined currents. Are such high voltages necessary, as one normally considers transistors to be low voltage devices?

Mr. Pollard has already raised the question of checking the multiplications. It is stated in the paper that known specimen multiplications have been tested and found correct. There is no mention of any method of checking the results of multiplication, apart from the verification on the Ferranti computer. Early experience with a point-contact transistor digital computer at A.E.R.E. has shown that the machine may be liable to make errors with certain digit patterns. It is difficult to predict which digit patterns may give rise to this sort of error. In our experience the reliability of the transistor circuits can be established only by use of a self-checking programme, involving a wide variety of digit patterns, such as a series of pseudo-random numbers generated by well-known arithmetical processes.

Mr. K. W. Cattermole: Like the authors, I have been working for several years on point-contact transistor circuits for pulse applications, although on a somewhat different basis. I have come to several of the same conclusions as the authors, although I am going to dissent from them in a moment because I do not follow them all the way.

The first point of interest is that, when one examines Dr. Chaplin's present circuits, one finds that they have departed a long way from the circuits in the classical Williams and Chaplin paper* of a few years ago. It is clear that they have been forced to make changes by two of the three main defects of the point-contact transistor. These defects are, first, the rather long reset time due to hole storage, and secondly, the great variability of the collector standing current i_{co} which makes bi-stable circuits rather difficult.

Having reached the same conclusion, I have gone a little further than the authors in the use of reactance resetting. They have an automatic reset worked by a somewhat ill-defined time-constant due to an inductance and certain properties of the transistor. It is my experience that one can do better by having a defined time built into the reactance in the form of a tuned circuit or rudimentary delay line. This enables one to reduce the number of clock pulse sources. I wonder what the authors feel about this.

There is one snag in trying to reset a point-contact transistor faster than its natural reset time: if this is done forcibly, the voltage of the collector may change while it is still conducting and cause a very large instantaneous dissipation. After some trials I came to the conclusion that it was not worth while. The transistors would not stand up to it as they would to the rather more gentle circuits.

Again, I should like to know the authors' views and whether they consider that their particular type of resetting deteriorates the transistor life as compared with the natural resetting by letting it revert when the hole storage has run out completely.

I suggest that unreliability is still the biggest defect in the point-contact transistor. I use the device myself for the same reasons as the authors, namely cost, speed and circuit versatility. But I have had apparatus running with point-contact transistors for a much longer period. The authors analyse their daily operation for about seven months and assure me that this represents something like eight hours a day five days a week. This comes to about 1 200 hours. I have had some pieces of transistor apparatus running for something of the order of ten times this duration and with more than ten times the proportion of failures. We cannot be dogmatic yet about long-term reliability. These were, of course, early model transistors, and they are always being improved.

Mr. G. W. Thompson: The authors, in their demonstration, showed how one can have intermediate access along such a storage device as a delay line. Normal practice is to use this intermediate access point as a "read" head, but here it has been used as a "write" head, and I must admit that I was not very clear as to its intention. It seemed to me that what the authors were advocating was that by this double injection of the "write" signal, attenuation could be reduced along the delay line. Unfortunately this seems to be at the expense of reducing the actual number of storage elements along the delay line, because if one introduced two signals, each identical in intelligence, one could achieve the same result by simply reducing the length of the delay line by half. Perhaps the authors would make some comment on this aspect.

The authors point out that the magnetostriction delay line has a number of advantages over other forms of storage system. One such advantage, they say, is speed of operation. The magnetic drum, however, is quite capable of operating at speeds of 100 kc/s; not only that, but it also offers access facilities at intermediate points. Furthermore, with a magnetic drum one can provide synchronization tracks which will maintain the associated clock and control circuits in synchronism with the rotating storage system.

With regard to computer processes involving multiplication, no means for checking would seem to be provided in the system described in the paper, and I am wondering whether consideration has ever been given to working back from the answer so that by dividing by one original product one could check that the result should be the other product, and vice versa.

Mr. C. J. Dakin: Does the material of which the delay line is made have to be nickel, or are other materials suitable? Secondly, is the material stressed mechanically in any way and is the stress, if there is one, at all critical? Lastly, why does it have to be stranded?

Dr. G. B. B. Chaplin and **Messrs. R. E. Hayes** and **A. R. Owens** (*in reply*): The discussion seems to be largely centred on the relative merits of point-contact and junction transistors. Mr. Cooke-Yarborough not only feels that the point-contact is faster and more versatile than the junction, but points out that its performance is already as good as the theoretical limit of the simple junction. Further weight is given to this view by Mr. Scott's intimation of even better things to come in the line of point-contact transistors.

On the other hand, Mr. Pollard and, with reservations, Dr. Grimsdale feel that the junction transistor is more reliable, and are prepared to wait for improved frequency characteristics.

We feel that point-contact circuits can be designed to be reliable at the present time, and their many other advantages make them very suitable for computer circuits. In general, when comparing point-contacts and junctions, similar circuit conditions and techniques should be assumed, to obtain useful results. Thus, for example, comparison should be made between circuits having approximately the same ratio of output to input pulse power. Under these conditions the rates of "turn on" and "turn off" of a point-contact are about ten times and twice that of a junction respectively. These are the figures which ultimately determine the relative maximum operating speeds of circuits. It is the "turn off" time which limits the point-contact to being only two or three times as fast as a junction, but if bottoming is prevented the point-contact becomes faster by a factor of ten in both respects, and operating frequencies in the megacycle region are easily obtainable.

Several speakers advocated caution in the use of base coils.

* Williams, F. C., and Chaplin, G. B. B.: "A Method of Designing Transistor Trigger Circuits," *Proceedings I.E.E.*, Paper No. 1428 R, January, 1953 (100, Part III, p. 228).

In particular Mr. Cooke-Yarborough contends that the circuit of Fig. 4, which includes a base coil, has little advantage over that of Fig. 2, which does not. The choice of base current, I_b, in Fig. 2 is limited to a range above and below which the circuit becomes unstable. Within this range the choice is decided by whether the circuit is desired to be sensitive to triggering or to resetting. The overriding consideration is that of sensitive triggering, and so I_b is chosen to have its minimum value. The use of a base coil, however, as in Fig. 4, gives an effective I_b when triggering of less than 1 mA, which is the current flowing in the damping resistor, and according to Mr. Cooke-Yarborough's figures, an effective I_b when resetting of 5 mA.

The coil has thus increased the sensitivity to triggering by as much as a factor of three and to resetting by approximately a factor of two. These are very significant factors, and the latter is based on a much more pessimistic estimate of the allowable ratio of I_c/I_b for stability than is normally assumed. Not only is the reset current now smaller, but its duration can be substantially decreased, since once the reset process has got under way the tendency of the coil to keep base current flowing completes the process even if the reset pulse is withdrawn. The total energy required from the reset pulse is in fact reduced to about a quarter of that in Fig. 2. An equally important reason for using base coils is stressed by Mr. Scott and Mr. Cattermole—namely that the magnitude and variation of i_{co} now present no problem.

Mr. Cattermole's attitude to base coils is that we have departed from the principles set out in the paper by Williams and Chaplin.* Of the many new techniques proposed in that paper, probably the two most useful for computing circuits were the earthed-emitter circuit of Fig. 6, which had a defined base current, and that of Fig. 20, which had a delay line in the base. The latter is apparently the circuit advocated by Mr. Cattermole, and we have merely simplified it by reducing the delay line to a single coil.

We agree with Mr. Cattermole that forcible resetting may lead to high collector dissipation and hence unreliability, but the question does not arise in the present case, since, in general, all collector currents are limited to about 15 mA, which is therefore the maximum resetting current.

Several speakers pointed out the desirability of some form of self-checking programme to supplement our reliability figures. This was in fact our intention, but circumstances prevented the work from being carried out.

Mr. Thompson queries the use of intermediate access in the magnetostrictive delay lines. We did not perhaps make sufficiently clear the purpose of the intermediate sender when demonstrating the storage system. Numbers injected via this sender are advanced in time with respect to the other numbers, and this property is made use of for setting up numbers and for progressively shifting them during multiplication. The fact that doubl amplitude pulses may at times be present in the line is or incidental but has to be allowed for in the design of the amplifie

Mr. Pollard sees a useful future for the magnetostrictive del line, whereas Mr. Thompson feels that the magnetic drum h many advantages. The main reasons for our choice of magnet strictive lines were their extreme cheapness, the absence moving parts and the large signal/noise ratio. In addition, the lines can be operated up to at least 1 Mc/s. In reply to D Stephen, drift, both in characteristics of the line and in oscillat frequency, is negligible compared with the $\pm\frac{1}{2}$ digit peri tolerance which is allowed. The problem becomes more acu as the number of stored digits increases, but simple precautio allow many hundreds of digits to be stored in one line. T width of the sending pulse will vary slightly with change transistor, but this has no effect on the timing of the receiv pulses and hence on the strobing. It does, however, have a sma but negligible effect on the amplitude of the received pulse. T physical length of the 64 digit line is 8 ft, and, in reply to D Grimsdale, the delay in the amplifier is about 3 microsec.

In reply to Dr. Dakin, nickel was chosen as the delay-li element because it has the highest magnetostrictive constant readily available materials. The effect of mechanical stress the mounting of the line would be to bias it in such a way as reduce the magnetostrictive constant. Owing to the skin effe at high frequencies it is desirable to make the ratio of surfa area to volume as large as possible, and this can be achieved using stranded wire, or alternatively thin-wall tubing.

We welcome Mr. Scott's encouraging estimate of the prese cost ratio of transistor to diode as 3 : 1.

Mr. Scott also raises the question whether stray capacitan adversely affects the circuits. The output impedance of t point-contact trigger circuit is virtually the forward impedan of the series emitter diode. This enables normal stray capac tances to be entirely neglected. A related problem which ma present difficulty in a large computer is that of the self-inductan of the wiring, since the circuits operate at relatively low impedanc

We are glad of support from Mr. Pollard in our policy deriving the clock system from valves since it allows the transisto to be underrun, which, as Mr. Scott points out, should increa transistor life.

Dr. Stephen feels that the ± 50 volt supplies which define t currents in the multiplier could be reduced. These voltag should be large compared with the waveform amplitude 14 volts, and reducing them would reduce the margin of safe with which the circuits have been designed. The overridi consideration in computer design is reliability, and we feel th the small increase in power consumption is well worth while.

May we end with a plea for the continued manufacture point-contact transistors until such time as junction units havi similar characteristics are commercially available.

* WILLIAMS, F. C., and CHAPLIN, G. B. B.: "A Method of Designing Transistor Trigger Circuits," *Proceedings I.E.E.*, Paper No. 1428 R, January, 1953 (**100**, Part III, p. 228).

537.211 : 621.315.59 : 621.838.3

Paper No. 2007 M
Mar. 1956

SOME THEORETICAL AND PRACTICAL CONSIDERATIONS OF THE JOHNSEN–RAHBEK EFFECT

By Miss AUDREY D. STUCKES, M.A., A.Inst.P.

(*The paper was first received 9th September, and in revised form 14th November*, 1955.)

SUMMARY

A theory of electrostatic attraction, influenced by field emission, is proposed to explain the Johnsen–Rahbek effect between the flat and polished surfaces of a metal and a semi-conductor in contact. The application and advantages of the effect as a clutch are described, and the results of a large number of life tests are summarized. It is concluded that the unreliability of such a clutch, due mainly to different amounts of wear, is caused by relatively small differences in the polished surfaces. Consequently it is suggested that the electrostatic clutch is not, in general, a practical proposition, but more likely applications of the effect lie in relay devices and valves for controlling gas flow.

LIST OF PRINCIPAL SYMBOLS

σ = Electrical conductivity, mho/cm.
ρ = Electrical resistivity, ohm-cm.
a = Radius of contact, cm.
R_c = Constriction resistance, ohms.
R = Resistance of semi-conductor, ohms.
C = Capacitance, farads.
W = Activation energy, eV.
V = Voltage, volts.
V_0 = Total applied voltage, volts.
F = Force, grammes.
E = Electric field, volts/cm.
A = Area, cm^2.
d = Gap width, cm.
T = Temperature, °K.
K = Boltzmann's constant.

(1) INTRODUCTION

When a flat and polished surface of a high-resistance semi-conductor is placed in contact with a flat and polished metal surface and a voltage is applied between them, a force is developed which tends to hold them together. This has become known as the Johnsen–Rahbek[1] effect, after two Danish scientists who observed that, when 400 volts was applied between plane and polished plates of lithographic stone and metal, a force of about 50 g/cm^2 was developed between them, which disappeared on removal of the voltage. In a systematic investigation they found that the force could be obtained with a number of minerals, including agate, limestone and marble, and also with some organic materials such as ivory, horn, bone and cellophane. The presence of moisture was essential, for only very small forces were obtained when the specimens were dried. The magnitude of the force was also found to vary with the electrical conductivity of the material, the degree of polish that could be given to its surface and the direction of the current flow. When the semi-conductor was negative with respect to the metal the maximum force was obtained; with the polarity reversed a force was developed instantaneously but gradually diminished, becoming negligible after a few minutes.

The subject was also investigated by others, including Rottgardt,[2] Bergmann[3] and Waszik,[4] mainly using naturally occurring materials such as marble, gypsum and slate—porous bodies whose conductivity depends on the presence of moisture in the pores (and on the direction of current flow) and is presumably ionic. Thus their conductivity (and the force developed) varies with atmospheric conditions and can be changed by the current flowing, since this tends to cause evaporation of some of the absorbed moisture. Moreover, the materials can gradually be decomposed by electrolytic processes. Consequently, although practical possibilities of the Johnsen–Rahbek effect in telephony and telegraphy were realized, the limitations imposed by the behaviour of ionic semi-conductors were probably responsible for the gradual lessening of interest in the various applications.

With the development of high-resistivity electronic semi-conductors (the resistance of which is independent of the direction of current flow) which do not rely on the presence of moisture (although moisture may affect them), it has become possible to make a hard, dense, ceramic material of controlled resistivity, capable of being ground and polished to optical limits.

Balakrishnan,[5] using sintered magnesium orthotitanate, Mg_2TiO_4 (a typical polycrystalline semi-conductor), with its surface and that of the metal each flat to within a few microns, was able to obtain a force of about 250 g/cm^2 with only 100 volts applied. Moreover, the force of attraction was not diminished by thoroughly drying the materials.

It seemed likely from these results that an improved semi-conductor capable of being polished to even finer limits would result in still larger forces. Discs of magnesium titanate with an apparent porosity of less than 2% have been ground and polished flat to within 0·25 micron or less on one side, the other side being covered with a sprayed or an evaporated metal layer to make contact over the whole face. When in contact with metal discs, similarly flat, a force of about 1 kg/cm^2 is developed with 100 volts applied and about 3 kg/cm^2 with 400 volts, almost irrespective of polarity; slightly higher forces are obtained with the metal negative.

The control of such forces by a small amount of electric power (e.g. 50 μA at 400 volts) seemed to offer the possibility of various useful applications. The investigations described in the paper were stimulated by these possibilities and by the need to understand the effect further.

(2) EXPLANATION OF THE JOHNSEN–RAHBEK EFFECT

The early investigators demonstrated that there was a contact resistance between the semi-conductor and metal which was large compared with the resistance of the semi-conductor itself, with the result that the greater part of the applied voltage was dropped near the surface contact. They also suggested that, since all the area of the surfaces was not in contact but separated by a thin air film, this greater part of the applied voltage would occur across the air-gap and hence lead to an electrostatic force of attraction across it. They were unable to explain the high contact resistance, except to consider it a property of the ionic semi-conductor.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Miss Stuckes is in the Research Department, Metropolitan-Vickers Electrical Co., Ltd.

A contact resistance will arise from the constriction of the current flow through the small regions of contact between the semi-conductor and the metal. When two surfaces are placed together the area in contact is extremely small, even when they are flat to the finest optical limits, for such surfaces will consist of hills and valleys which cannot be compressed into uniform contact except by enormous forces. Bowden and Tabor[6] have shown that two steel surfaces, 21 cm^2 in area, flat to within about a micron and placed together under a load of 5 kg, touch at only about five points, each with a contact radius of approximately 10^{-2} cm, i.e. with only 1/40 000 of their total area in contact.

It is well known that the constriction of current flow between a metal and a semi-conductor of relatively low, uniform conductivity, through a small contact, gives rise to a resistance $R_c = 1/4a\sigma$, where the electrical conductivity of the metal is regarded as infinite. Thus, when a voltage V_0 is applied between the flat and polished metal and the semi-conductor, this constriction offers the major part of the electrical resistance. The surface of the metal can be regarded as an equipotential, whereas over the semi-conductor surface there is a rapid change in potential in the neighbourhood of each contact point (Fig. 1). This leaves

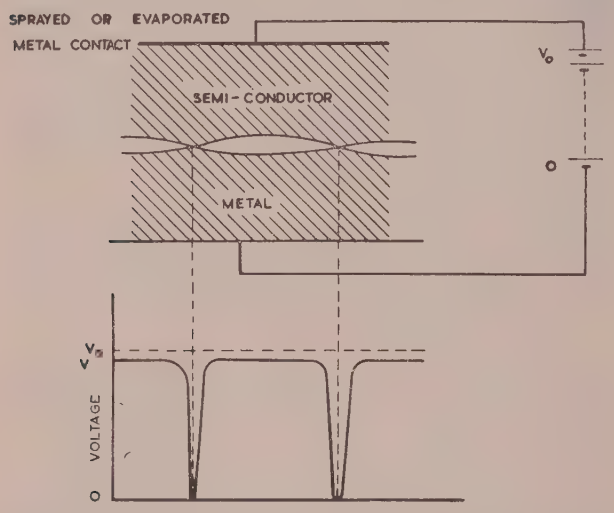


Fig. 1.—Potential distribution at surface of semi-conductor.

much of the semi-conductor surface at a potential V, which is variable, but only a little less than the total applied potential V_0.

This potential difference across the very small gap between the metal and the semi-conductor surfaces gives rise to an electrostatic force between them. In the ideal case, where two surfaces are parallel and separated by a distance d, except for one or two negligible regions of contact, the force is given by

$$F = \frac{AV^2}{8\pi d^2} = \frac{AE^2}{8\pi} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It is therefore expected that the force should be directly proportional to the square of the applied voltage, since this is only a little greater than the voltage in the gap.

On this basis the Johnsen–Rahbek effect may be represented by the equivalent circuit shown in Fig. 2(a), where the constriction resistance, R_c, is expected to be very much greater than the resistance of the semi-conductor, R; C represents the capacitance across the gap between the surfaces. Such a circuit will have a linear voltage/current characteristic whose slope is given by $R_c + R$.

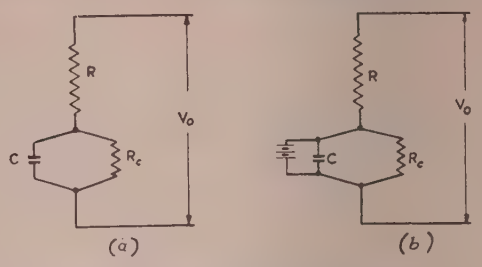


Fig. 2.—Equivalent circuit of the Johnsen–Rahbek effect.

(2.1) Factors influencing the Johnsen–Rahbek Effect

The observations of most investigators have indicated that the force produced is not always proportional to the square of the applied voltage. Rottgardt[2] found it to be proportional to higher powers of the voltage, Balakrishnan[5] found an approximate square law to hold, and the author has found that the force/voltage relationship (Fig. 3) is a steadily varying function, the force ultimately becoming constant.

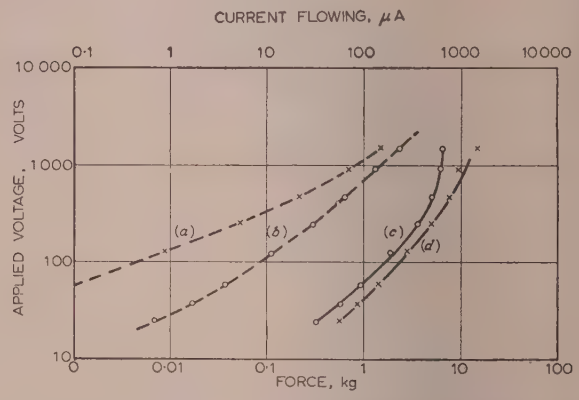


Fig. 3.—Variation of force and current with applied voltage for steel in contact with a disc of magnesium titanate, of 2 cm^2 area and 6·8 megohms resistance.

The surfaces were concave to about 1/4 micron
(a) Current/voltage, metal negative.
(b) Current/voltage, metal positive.
(c) Force/voltage, metal positive.
(d) Force/voltage, metal negative.

Simple argument leading to a square law does not consider the effect of two properties of the electronic semi-conductor, namely its negative temperature coefficient of resistance and its tendency to have surface barrier layers of high resistivity; it also ignores the fact that field emission will occur above certain field strengths across the gap.

The negative temperature coefficient of resistance will cause the current to increase more rapidly than the voltage when the heat generated at the regions of contact causes an appreciable rise of temperature there. The constriction resistance of the contacts would then decrease relatively to the resistance of the semi-conductor as a whole, and thereby reduce the proportion of voltage across the gap; this would tend to lessen the rate of increase of force with applied voltage. The effect of this heating on the voltage/current characteristic has been calculated by the author[7] for a semi-conductor obeying the approximate law $\rho = \rho_0 \varepsilon^{-W/2kT}$. From such calculations it seems certain that with the semi-conductors used in the present experiments (which have a resistivity of 10^6 ohm-cm or more at room temperature, and an activation energy of about 0·86 eV) heating could not

account for any noticeable departure from Ohm's law below 200 or 300 volts.

Barrier layers cause rectification and departures from Ohm's law at voltages above 0·1 volt. With most materials such barriers break down at some voltage less than, say, 50 volts; the exceptions are semi-conductors with a very low density of donor or acceptor centres (and hence a thick space-charge barrier layer), and with sufficiently high carrier mobility to conduct effectively with the low density of carriers. With magnesium titanate, however, such rectification is not usually very noticeable, the density of donor centres required to produce conductivities of 10^{-6} ohm-cm is high and the mobility of carriers too low to be measured without great difficulty.

The third and perhaps most important factor in the Johnsen–Rahbek effect is the occurrence of field emission across the gap between the metal and the semi-conductor. The mean free path of electrons at atmospheric pressure is of the order of 0·25 micron, so that they will suffer few collisions in crossing a gap of similar width. Field emission[8] from "cleaned-up" surfaces is known to become appreciable at field strengths in the region of 10^7 volts/cm, depending upon the work function of the material from which the electrons are being emitted. From surfaces exposed to normal atmospheric contamination it is likely that field emission will start at lower field strengths. Since a very small change of field strength produces a very large change of current, the field strength which can be applied to a surface is virtually limited to that at which appreciable emission begins. Since the gap is not uniform, emission first occurs around the contact points, where the field is greatest. Increase of applied voltage increases the area over which emission is taking place until it covers the whole surface; still further rise will augment the current but leave the field in the gap almost constant.

With this picture one would expect that the force between the surfaces would increase as the square of the applied voltage only at rather low voltages, and less rapidly as the area of emission spreads, finally remaining independent of further increase when the voltage is large. In the lowest range of current Ohm's law would be followed, the effective resistance being $R_c + R$, whereas in the highest range of current, when emission covers the whole area, the effective resistance becomes R and the current/voltage relationship becomes linear with an intercept on the voltage axis roughly equal to the voltage required to produce emission field-strength across the average gap. In the intermediate range, according to Sillars,[9] the current around a contact area should be proportional to $V^{3/2}$ or V^2 for convex surfaces, and to higher powers of voltage for concave surfaces, the exact relation depending upon the configuration.

Since the effective work function of a semi-conductor is almost certainly less than that of a metal (so that the field strength required to produce emission would also be less), it might also be expected that the voltage and hence the force would be lower for a given current when the semi-conductor was negative.

A fourth factor[2] that might influence the effect is that the force may be sufficient to pull the surfaces closer together, presumably either by deforming the contact areas or by crushing the points of contact. This would result in the force increasing more rapidly than the square of the applied voltage.

(2.2) Measurements of Current, Force and Voltage

Measurements were taken on discs of magnesium orthotitanate 1·6 cm in diameter in contact with steel, each polished surface being concave by approximately 0·25 micron. One member was rigidly attached to the bench, the other, laid on top of it, could be attached to a spring balance to measure the force required to pull it off.

In one set of measurements the voltage/current curves were taken without disturbing the discs; a typical pair is shown in Fig. 4. These tend to become linear at high voltages, with a slope resistance of a few megohms, and are tending to become linear at low voltages, with a slope resistance about a hundred times greater. In the middle range the current varies as the square or a higher power of the voltage, and is higher when the semi-conductor is negative than when the metal is negative.

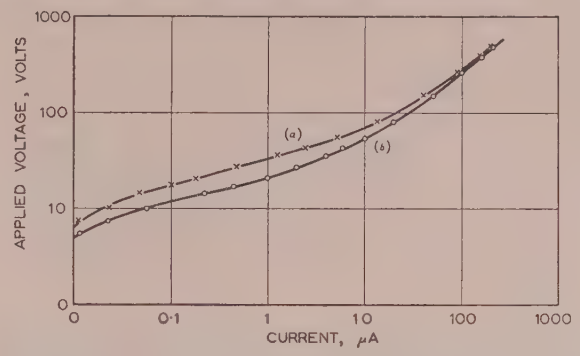


Fig. 4.—Variation of current with applied voltage for steel in contact with a disc of magnesium titanate, of 2 cm² area and 1·8 megohms resistance.
The surfaces were concave to about 1/4 micron.
(*a*) Metal negative.
(*b*) Metal positive.

In another set of measurements the force required to separate the discs was measured, in addition to the voltage and current. Since the current decreased slightly as the pull on the upper disc was applied, the value recorded was the initial one before the pull was applied. After each separation the upper disc was replaced as nearly as possible in the same position, and each observation was repeated several times. A typical plot of the average readings is shown in Fig. 3. The voltage/current curves have features similar to those observed in Fig. 4, but were not taken to voltages low enough to observe any trend towards linearity at the lower end. The attractive force is lower when the semi-conductor is negative than when the steel is negative, and is roughly proportional to the 3/2 power of the voltage in the low range, becoming constant (more definitely when the semi-conductor is negative) in the high range.

After these observations the polished surface of the semi-conductor was provided with a metal film in the same way as the other side and its resistance was measured; the values found are recorded in Figs. 3 and 4.

(2.3) Discussion of Results

The voltage/current curves cannot be satisfactorily explained by heating of the contact regions, because the theory of this effect[7] indicates that they should remain linear up to at least 200–300 volts. Nor are they in any way typical of barrier-layer effects, because rectification is not very marked even at 5–10 volts, where it would be expected to be greatest.

Both the voltage/current and the voltage/force curves are, however, roughly consistent with the behaviour to be expected when field emission develops between the surfaces, as discussed in Section 2.1. The numerical values are of the order expected; they do not bear close quantitative comparison, but the arguments outlined have taken little account of the large-scale variations of gap-width, and none at all of small-scale irregularities or of possible local variations in work function. The limiting value of pull when the semi-conductor is negative corresponds to a force of 3 kg/cm², which would be produced by a field of 2·6 MV/cm. If the maximum gap is assumed to be 5×10^{-5} cm, this field strength would require 130 volts across the gap, leaving

about 1 kV to pass the current of 100 μA or so through the disc, which attributes to the disc a resistance of about 10 megohms, compared with the measured value of 6·8 megohms. The highest pull recorded in Fig. 3 corresponds with a force of 7 kg/cm^2, accompanied by an average current density of 70 μA/cm^2, the steel being negative. This force would be produced by a field of 4 MV/cm, and the work function calculated by inserting these values in the Fowler–Nordheim[8] equation would be about 1·3 eV. This is undoubtedly lower than the value for a clean metal surface, but may be roughly appropriate to a surface exposed to adsorption of gases and vapours from the atmosphere.

Deformation of the contact areas doubtless occurs, but would not by itself explain either the voltage/current or the voltage/force curve. It would, however, explain the slight change in current when a pull is applied.

The preceding comments apply primarily to the experiments recorded here. Those by Balakrishnan were on discs concave by one micron or more, so that forces were less and field emission would be much less important for the same voltage. An attractive force varying as the square of the voltage (or even higher power of voltage if deformation of the contacts occurred) would be expected, as was, in fact, observed. Similar remarks may apply to the earlier investigations; Rottgardt was the only worker to estimate the size of his gap, and gave it as 7 microns.

(2.4) The Equivalent Circuit of the Johnsen–Rahbek Effect and its Time-Constant

It may be concluded from the experimental results that the Johnsen–Rahbek force is an electrostatic force as explained in Section 2, but its characteristics are modified by the occurrence of field emission. With these considerations in mind the equivalent circuit shown in Fig. 2(*a*) remains unchanged, but the resistance R_c will now be dependent on the supply voltage. At low voltages R_c will represent the constriction resistance of the contact points; it will be constant, very much larger than the resistance R of the semi-conductor, and will depend only on the number and size of the contacts. Under these conditions the voltage, V, across the gap at any time t after a voltage V_0 is applied is given by

$$V = \frac{V_0 R_c}{R + R_c}\left[1 - \exp - \frac{(R + R_c)}{CRR_c}t\right] \quad . \quad . \quad (2)$$

Initially, therefore, the voltage build-up is controlled by a time-constant CR, since $R_c \gg R$, but as the supply voltage increases and field emission begins, the effective resistance R_c will decrease and will ultimately be replaced by a limiting voltage when emission is occurring over the whole area. When this has happened the equivalent circuit is that shown in Fig. 2(*b*), where C and R_c are in parallel with a constant-voltage supply. This circuit has a zero time-constant, so that a small change of V_0 would be accompanied by a corresponding change of current, but the current would cease to change immediately V_0 ceased to change.

It is difficult to analyse precisely the whole sequence of events when V_0 is suddenly changed from zero to a value which will ultimately cause emission over the whole surface. At first eqn. (2) applies and V grows exponentially, but as field emission spreads one may think of portions of the capacitance C being short-circuited by a constant-voltage supply, so that a diminishing capacitance is being fed through R and the "time-constant" drops steadily to zero. The process is not exponential, but its time-scale is still controlled by CR, with the qualification that the lower the voltage (relative to V_0) at which field emission occurs across the gap (i.e. the smaller the gap), the earlier is the exponential growth superseded and the steady state established. Thus the increase of the capacitance, and therefore of the simple time-constant RC caused by a reduction in the gap, is to some degree compensated by the also diminishing time before emission begins. It also follows that the higher the applied voltage V_0 the quicker the stage of emission will be reached and therefore the sooner the force will be developed.

In many applications of the Johnsen–Rahbek effect this time delay must be considered, and the time-constant is the simplest index of it. If the surfaces are each concave to within 0·25 micron the capacitance will be about 2×10^{-3} μF/cm^2. Time-constants of the order of 1 millisec must therefore be expected.

Owing to the inherent time delay there is a limit to the frequency of an alternating voltage supply to which the effect will respond. To obtain forces of an order similar to those obtained with d.c. supplies the time-constant should not be greater than one-quarter of a cycle, so that if 1 millisec is the minimum time-constant, reasonable forces will be obtained only up to frequencies of about 250 c/s. Some observations on the type of clutch described in Section 3.1, using a variable-frequency supply, roughly confirmed these conclusions.

(3) APPLICATIONS OF THE JOHNSEN–RAHBEK EFFECT

Although many of the original ideas for applying the Johnsen–Rahbek effect have been superseded, others have become possible because of the greater forces obtainable with the superior electronic semi-conductors.

One obvious application would be an electrostatic clutch with the two flat, polished surfaces of the metal and the semi-conductor held lightly in contact; clutching and declutching would take place on application and removal of a voltage between them. It would be expected to have a more rapid time response, lower power consumption and smaller physical dimensions than existing types of clutch dealing with comparable forces. The possibilities of such a device and of some relay-type applications have been investigated in this laboratory.

In all applications of the effect it must be realized, however, that unlike a magnetic force the electrostatic attraction is not noticeable until there is almost physical contact between the polished surfaces. Since the force is inversely proportional to the square of the distance separating the surfaces, it will be extremely small if they meet at an angle, so that it is unsatisfactory to rely on the electrostatic attraction to pull the surfaces into closer contact and thereby develop a greater force between them.

It is also necessary either to connect a bleed resistor across the device or to design a circuit that connects the metal and semi-conductor together when the voltage is removed, otherwise the capacitance between the surfaces can discharge only through the constriction resistance of the contact points, giving a much longer time-constant for the decay of the electrostatic force than for its build-up.

(3.1) An Electrostatic Clutch

An essential feature of a clutch operating by virtue of the Johnsen–Rahbek effect is that the two clutch faces are either always in intimate contact or are brought into such contact before the application of the voltage. Since it is impractical to mount the two clutch shafts in their bearings with sufficient accuracy to maintain the polished surfaces parallel within optical limits whatever their relative orientation, one member must be made to align itself. A section through a typical clutch is shown in Fig. 5.

Only one semi-conductor material has been tried, namely magnesium orthotitanate. The best results have been obtained using a hard steel for the metal, usually nitrided nitralloy, but

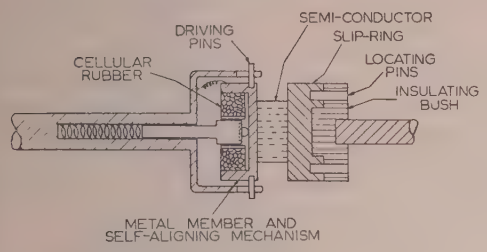


Fig. 5.—Section through an electrostatic clutch.

even in this case the hardness of the titanate was the greater. A clutch incorporating two semi-conductors instead of the usual metal and semi-conductor was tried, but this usually had a very short life.

Preliminary experiments on the clutch indicated that the type of load has a considerable influence on the wear of the surfaces. The operating life was shortened when the load applied was independent of speed (e.g. by a friction brake) compared with an equal load of similar inertia applied gradually (e.g. by an eddy-current dynamometer). It was also shortened if the inertia was increased without any increase in the load.

Some indication of the magnitude of the loads that can be handled by this device is given by the torque measurements on a clutch 1·6 cm in diameter. The results given in Fig. 6 indicate

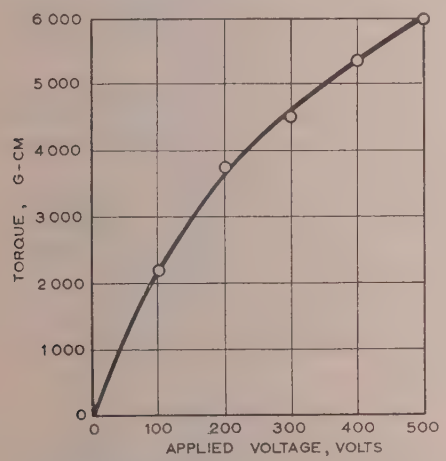


Fig. 6.—Static torque to produce slip plotted against applied voltage for clutch members 1·6 cm in diameter.

the maximum transmissible torque at any voltage from 100–500 volts, and those in Fig. 7 give the pick-up torque, i.e. the torque transmitted at various relative speeds of slip from 100 to 1 100 r.p.m. The torque transmitted will be proportional to the cube of the diameter, and some experiments on discs 3·1 and 3·8 cm in diameter gave results consistent with this. However, the wear was so rapid, especially on the larger discs, that this work was not pursued.

In general, when the clutch breaks down and fails to drive a given load after repeated operations, a small amount of powder, estimated to be only a few microgrammes, is found on the polished surfaces. After cleaning, the clutch will once more drive the given load, although the life after cleaning does not appear to be related to its life prior to breakdown. In almost every case it has been shown by spectrographic analysis that this powder is the result of wear of the metal surface.

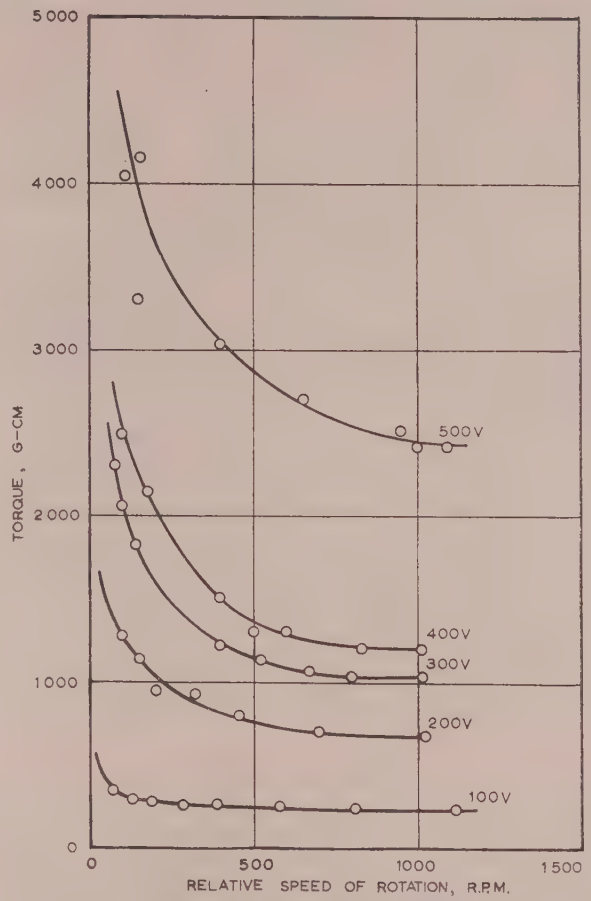


Fig. 7.—Torque transmitted when clutch members (1·6 cm diameter) are slipping.

(3.1.1) **Life Tests.**

It seemed likely that in general this type of clutch would be suitable for on-off applications, where the surfaces are normally locked together during the transmission of torque, the clutch being employed only occasionally to accelerate (or decelerate) the driven shaft. If used as a slipping clutch to provide smooth output control by control of voltage in such a way that the surfaces rarely lock together, then prolonged heating by slipping is likely, not only to cause increased wear of the polished surfaces, but also to lead to electrical instability as a result of the large negative temperature coefficient of resistance of the semi-conductor.

Tests have therefore been carried out using a number of similar clutch pairs (magnesium titanate and nitrided nitralloy discs 1·6 cm in diameter) to determine how long they would continue to transmit a reasonable torque during repeated on-off operations. A standard load of 1 000 g-cm at 1 000 r.p.m. provided by an eddy-current dynamometer was chosen, and the clutch was energized at 500 volts at a rate of 720 times per hour with approximately equal on-off periods.

The most striking feature of several hundred life tests carried out under similar conditions was the inconsistency: a variation in life from a few hundred to a hundred thousand operations was obtained.

Variations in the physical properties of the semi-conductor, which can easily occur from disc to disc owing to the nature of the manufacturing techniques involved, were first considered as

a likely cause of these inconsistencies. However, no correlation could be found with porosity, grain size, presence of impurities, surface roughness or hardness.

Since the process of wear occurs during the time of slipping with the clutch energized, it seems likely that the variations in life must be due to variations in this pick-up time or in the force acting during that time. Both of these factors could have varied in the tests, since neither the resistance of the semi-conductor nor its surface finish was strictly controlled.

From a simple analysis of the picking-up process it has been estimated that the variations in the electrical time-constant and pick-up torque that occurred could account for a variation in wear by a factor of about five to one. However, the presence of even extremely small particles produced by wear will drastically alter the gap between the surfaces, since this initially is only half a micron at its greatest. Thus the pick-up torque will be progressively reduced and greater amounts of wear will occur with each operation. It also seems likely from the analysis that the rate of wear will become very rapid once the pick-up torque approaches the value of the load torque.

Consequently, it is suggested that to obtain consistent life from clutches operating on this principle it would be necessary, not only to control the materials and the obvious mechanical factors, but also to keep extremely close control of the resistance, and—what is more difficult—of the exact surface configuration of the metal and semi-conductor. There is no evidence, however, to suggest that even under these conditions a 1·6 cm-diameter clutch energized at 500 volts could consistently transmit a load of 1 000 g-cm for more than 100 000 operations.

(3.2) Relay-Type Applications

Applications of the Johnsen–Rahbek effect can be found which use the direct force of attraction between the surfaces rather than friction between them. Although this force cannot be used to bring the two surfaces together from some distance apart, if they are placed together by some means it can continue to hold them together. Thus, if one member is fixed and the other is periodically moved into contact with it by some mechanical, hydraulic, pneumatic or other device, the application of a suitable signal voltage between them will stop the periodic movement and removal of it will cause the movement to recommence. In this way a small amount of electrical power (say 1 mW) can exert a control action on other equipment. Some experiments on a device of this nature have shown that it continues to respond to a signal voltage for a few million operations without damage to the surfaces.

It has also been shown that the effect can be utilized in a similar fashion in the design of a valve to control rate of gas flow in accordance with electrical signals. The valve can consist simply of a semi-conductor and metal held together by spring pressure, one of its members having a central hole through which the gas to be controlled must pass. The gas will force the surfaces apart against the spring pressure, and for small flows the gap may be so small that the application of a suitable voltage will pull the surfaces together and prevent any further gas flow until the voltage is removed. With larger gas flows, where the separation becomes too great for the applied voltage to pull the surfaces together, one member must be made to oscillate, allowing a puff of gas to escape at each separation.

It is possible to arrange the two members so that they are part of an orifice-reservoir system which is self-oscillatory. In such a system, one member having a central hole is rigidly clamped to a small reservoir into which gas is fed either at a constant rate or through a high impedance. The other member, with its polished surface in contact with that of the fixed member, is mounted on a spring system designed to give a low natural frequency. If a small steady voltage is applied, such that the force between the surfaces is not great enough to withstand the full pressure available, the pressure in the reservoir builds up until the spring-mounted member is forced off. With suitable design the members remain apart long enough to allow the reservoir to empty and thereby allow the restoring spring to bring the surfaces into contact again, so that the cycle can repeat itself. This system allows a pulsating flow of gas, which can be shut off by application of a suitably large voltage and the surfaces then kept together against the total pressure available. Such a valve has been constructed using a 1·6 cm-diameter semi-conductor. It oscillates at a low audio frequency when 100 volts is applied and closes when 500 volts is applied; it can then withstand pressures of the order of 5 atm without difficulty.

(3.3) Limitations to Applications of the Johnsen–Rahbek Effect

It has been found that the current flowing through the metal/semi-conductor junction can itself ultimately cause deterioration of the polished surface of whichever member has positive polarity. The damage becomes obvious after a current of some tens of microamperes has been flowing continuously for 50 hours in a normal atmosphere. With a metal member of nitrided steel, rust formation occurs, and with the semi-conductor of magnesium titanate, its polish disappears. The damage, in general, is not uniform but occurs in a ring starting near the periphery and sometimes extends almost to the centre. However, if the members are kept in a normal type of calcium chloride desiccator, a similar current can be passed for over 500 hours without any sign of deterioration to either the semi-conductor or the metal, irrespective of polarity.

It seems likely that the area of damage denotes the region over which field emission was occurring, and is the result of some secondary action of electronic bombardment of the surface in which moisture plays an essential part.

This difficulty might be overcome by choosing a suitable corrosion-resisting metal and always applying the positive potential to it. Stainless steels have shown little signs of damage after 70 hours in a normal atmosphere. A more satisfactory solution for relay applications would be obtained by mounting the metal and semi-conductor in a sealed unit with a drying agent. Its use as a gas-control valve might have to be confined to dry systems.

(4) CONCLUSIONS

The Johnsen–Rahbek force can be explained as the electrostatic force across the gap which exists (except at a few spots) between two nominally flat, polished surfaces in contact. The force is proportional to the square of the applied voltage until the field in the gap reaches, locally, the critical value where field emission occurs. Further increase in the applied voltage cannot then increase this field, but merely extends the region over which it is acting, until ultimately it covers the whole area of the gap. The force must therefore increase more slowly with increase in applied voltage and must ultimately remain constant. Slightly higher forces are obtained when the metal is negative with respect to the semi-conductor, presumably because it has the higher work function and therefore the higher critical field for the onset of emission. Forces approaching 25% of the theoretical limit (calculated from the work function of a clean metal) have been attained.

A clutch operating by virtue of the electrostatic force will pick up and transmit a load exerting a torque of a few thousand gramme-centimetres when incorporating a semi-conductor 1·6 cm in diameter and energized at 500 volts. However, it has proved to be unreliable in a large series of life tests, and although some

clutches operated 100 000 times, others similarly prepared have operated only a few hundred times. It appears that large differences in the amount of wear may arise from comparatively small variations in the surface curvature or surface structure, so that it would be difficult to achieve reliable operation in practical conditions.

There may be more useful applications as relay devices and gas-control valves, with the proviso that dry conditions are probably necessary for reasonably long service.

(5) ACKNOWLEDGMENTS

The author wishes to thank Dr. R. W. Sillars for his continued interest and encouragement, Mr. D. P. Jones for many stimulating discussions, particularly with reference to the preparation and properties of the semi-conductor, and Mr. P. J. Heywood for much of the experimental work on the clutch. The author also wishes to thank Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish the paper.

(6) REFERENCES

(1) JOHNSEN, A., and RAHBEK, K.: "A Physical Phenomenon and its Applications to Telegraphy, Telephony, etc.," *Journal I.E.E.*, 1923, **61**, p. 713.

(2) ROTTGARDT, K.: "Die elektrische Anziehung nach Johnsen–Rahbek und ihre Anwendung in der Hoch frequenztechnik," *Jahrbuch der Drahtlooer Telegraphie und Telephone*, 1922, **19–20**, p. 299.

(3) BERGMAN, L.: "Über einige Messungen am elektrostatischen Relais," *Zeitschrift für Technische Physik*, 1923, **4**, p. 11.

(4) WASZIK, J.: "Beitrag zur Erklärung der elektrischen Anziehung, die als Johnsen–Rahbek–Phänomen bezeichnet wird," *Zeitschrift für Physik*, 1924, **5**, p. 29.

(5) BALAKRISHNAN, C.: "Johnsen–Rahbek Effect with an Electronic Semi-Conductor," *British Journal of Applied Physics*, 1950, **1**, p. 211.

(6) BOWDEN, F. P., and TABOR, D.: "The Friction and Lubrication of Solids" (Clarendon Press, Oxford, 1950), p. 30.

(7) STUCKES, A. D.: "Electrothermal Behaviour of Point Contacts to Semi-Conductors," *Proceedings of the Physical Society*, 1953, **66**, p. 570.

(8) FOWLER, R. H., and NORDHEIM, L.: "Electron Emission in Intense Electric Fields," *Proceedings of the Royal Society*, A, 1928, **119**, p. 173.

(9) SILLARS, R. W.: "The Effect of Field Emission on the Behaviour of Semi-Conductor Contacts," *Proceedings of the Physical Society*, B, 1955, **68**, p. 881.

621.384.6 : 621.319.3

Paper No. 1869 M
July 1955

AN ELECTROSTATIC PARTICLE ACCELERATOR

By D. R. CHICK, M.Sc., B.Sc., Member, and D. P. R. PETRIE, M.Sc.

(*The paper was first received 8th October*, 1954, *in revised form 28th January*, 1955, *and in final form 24th March*, 1955. *It was published in July*, 1955, *and was read before the* MEASUREMENT AND CONTROL SECTION *25th October*, 1955.)

SUMMARY

A pressurized electrostatic generator has been designed to give as nearly as possible a uniform field in the dielectric medium, and the expected performance is calculated assuming that the insulating gas breaks down at 200 kV/cm.

The construction of the generator and of a multi-section ion-accelerator tube is also described, and the performance of the machine both as a voltage generator and as an ion accelerator is reported. Novel methods of metering currents at the high-voltage electrode allow the behaviour of the generator in the presence of an ion beam to be analysed.

The voltage stabilization of the accelerator and the stability of the main servo loops are discussed.

A beam analyser employing electrostatic deflection provides precise absolute measurement of the beam energy. Calibration points on the high-voltage scale at 874·7 and 2567 kV have been accurately determined.

The peak performance is 3·8 MV with the 9 ft tube and 5·5 MV without the tube. Proton beams up to 3·25 MeV have been obtained with an energy resolution of 1 in 1 200.

(1) INTRODUCTION

A pressurized electrostatic generator of the charged-belt type was designed and constructed at Aldermaston as part of a programme of accurate measurements on the nuclear energy levels in the lighter elements. The aim is to provide an ion beam whose energy is stabilized and known to a high degree of precision. The generator potential is applied to an evacuated tube which is built within the generator itself, and along which the beam of ions is accelerated. A beam analyser, employing electrostatic deflection, is used to measure the ion energy absolutely by comparing it ultimately with a standard cell.

Since the generator is a constant-current source, its potential is controlled by varying the current load. This is easily done, with a fast response, by modulating an electron beam from a gun at the base of the tube. The electrons are accelerated in the same vacuum tube as the protons, and act as a load on the generator. The potential can be held constant to one part in 1 200 or better by means of servo loops between the voltmeter and analyser and the electron gun. This will be described in more detail later.

The generator itself is based on a design developed by Van de Graaff at the Massachusetts Institute of Technology, though it has many mechanical details in common with the generator at the Atomic Energy Research Establishment, Harwell, described by Fortescue and Hall.[1] Certain improvements in the electrostatic design have been made, notably the use of elliptical-section guard rings around the periphery of the generator column plates, instead of the usual circular roll, in order to reduce the radial field. The theory of these guard rings was developed by Boag[2] of the Medical Research Council, but they were not incorporated in the M.R.C. machine.

In the design of the accelerator tube, the most notable feature is the width of the tube, which is the same as that used by the M.R.C. but considerably greater than is commonly used. The internal diameter of the walls is 10 in and the electrode bore is 6 in. The wide tube provides faster pumping and lessens the chance of electrodes being struck by stray protons from the beam. Other novel features of the tube will be described later.

Considerable care has been taken to make the electrostatic design of the whole generator column as near to optimum as possible within a given sized pressure tank. The theoretical aspects will therefore be considered first in some detail, after which the construction and performance of the machine will be reported.

(2) GENERATOR DESIGN CONSIDERATIONS

The optimum performance for a generator of given size will be achieved by a design which allows the dielectric gas to be stressed as uniformly as possible through the volume. The axial length of the generator will, however, be determined by considerations of surface flashover along the supporting solid insulators and the belt, while, as we shall see, the presence of an ion-accelerating tube profoundly modifies the behaviour of the generator as a voltage source.

In considering the most efficient proportions of the electrodes and their detailed profiles it is assumed that breakdown will occur if a certain critical field strength E_m is exceeded. This is justified by the fact that the dimensions of the device are large compared with the mean free path of an electron in the gas, and the field strength is therefore almost constant over one free path (its rate of variation does not exceed 50% per cm); hence the energy gained by an electron in a free path, which is the primary factor governing breakdown, is simply proportional to the field strength at the point in question.

Several different electrostatic designs were compared using the above criterion, and an attempt was made to achieve a design in which the insulating gas was stressed nearly equally over the several critical electrode surfaces.

The generator column (Fig. 4) consists of a series of metal plates equally spaced, insulated from one another, and terminating in a spinning with cylindrical wall and a hemispherical cap. Each plate has around its periphery a smooth metal guard-ring roughly elliptical in section.

The greatest fields occur on the surface of the top terminal and are readily calculated to a first approximation by treating the terminal and tank as concentric spheres in the hemispherical part and as concentric cylinders elsewhere.

Since the field E decreases outwards in both the spherical and cylindrical gaps the average value of E and hence the potential V between top terminal and tank (i.e. $\int Edr$ over the gap) can be increased by the use of intermediate shields arranged in position and potential such that the electric field on the outside of each shield is made equal to E_m.

Boag has derived the values of successive shield radii to make V/E a maximum for a given value R of tank radius. He concludes that for most conditions the improvement in V/E is rather insensitive to the value of r_i/r_{i+1} (r_i being the radius of

Mr. Chick and Mr. Petrie are at the Research Laboratory of Associated Electrical Industries Ltd.

the ith intershield). Thus a choice of r_i/r_{i+1} = constant, or $(r_{i+1} - r_i)$ = constant will give breakdown improvements which are both very close to the maximum possible obtainable.

A useful way of demonstrating the effect of the intershields is to plot the ratio of the potential V_n between terminal and tank with n intershields to the maximum possible value V_∞ for an infinite number of intershields.

For concentric spheres it can be shown that

$$V_n/V_\infty = \sum_{i=0}^{n} q\frac{r_i}{r_0}\left(1 - \frac{r_i}{r_{i+1}}\right) \quad . \quad . \quad . \quad . \quad (1)$$

where r_0 is the radius of the terminal spinning, $R = r_{n+1}$ is the tank radius, and $q = r_0/(R - r_0)$, while for coaxial cylinders the expression is

$$V_n/V_\infty = \sum_{i=0}^{n}\left(q + \frac{i}{n+1}\right)\ln\frac{q(n+1)+(i+1)}{q(n+1)+i} \quad . \quad (2)$$

These expressions are plotted in Figs. 1 and 2 for various values of n, and it should be noted that the cylindrical geometry is

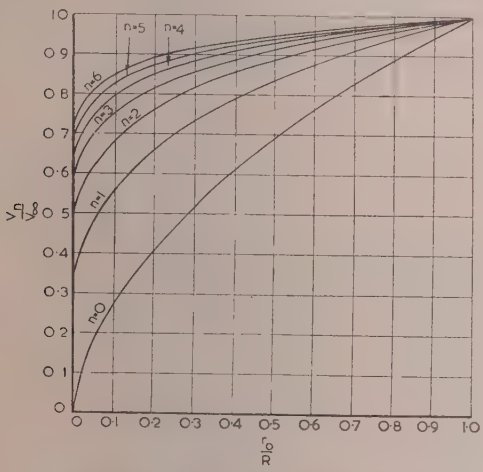


Fig. 1.—The effect of equally spaced concentric cylindrical shields.

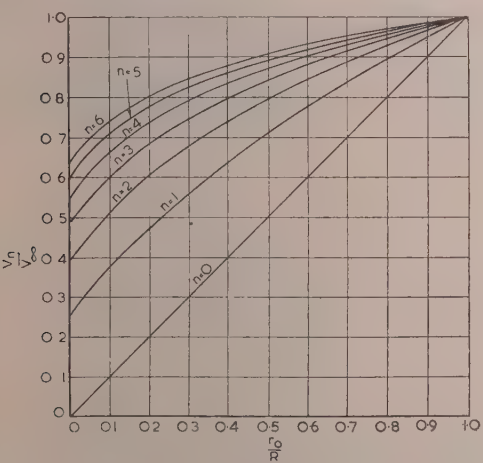


Fig. 2.—The effect of equally spaced spherical shields.

inherently stronger than the spherical, and that the gain by increasing the number of intershields decreases for each additional intershield. Considerations of cost, weight and difficulty of handling, together with the fact that the pressure vessel diameter for the Aldermaston generator was fixed at 70½ in by the existence of suitable tools, led to the choice of a design based on one intershield. Further careful consideration led to the choice of $r_0/R = 0{\cdot}5$, which gives an optimum performance when the intershield is not in use. It gives a terminal diameter of 35 in, and the maximum improvement in the spherical region due to the intershield is only 9% less than the best obtainable. This particular case was analysed in detail to give an intershield radius of $r_1 = 0{\cdot}733\ R$, and this radius gives

$$V/E_m R = 0{\cdot}352 \text{ for spherical symmetry,}$$
$$V/E_m R = 0{\cdot}41 \text{ for cylindrical symmetry.}$$

Thus the cylindrical part is stronger by the factor 1·17, and hence the broken nature of the cylindrical stack must not be allowed to cause a distortion factor greater than this.

Before the maximum electric fields occurring in any particular design can be calculated, three important field distortion factors must be estimated; first, the profile of the guard-rings, which modifies the maximum radial field in the cylindrical regions; secondly, the effect of eccentricity of the generator column with respect to the tank; thirdly, the design of the intershield profile, which modifies the field between the shield and the tank.

(2.1) The Guard-Ring Profiles

Since the mean radial field is considerably greater than the mean axial field, the best shape for the guard-ring cross-section is an ellipse-like curve presenting its minimum curvature to the radial field and its maximum curvature to the axial field.

In Section 13.1 analytic expressions are given for the shape of such curves and the corresponding field distortion factors, f_h to be applied to the radial field, and f_{ax} to be applied to the axial field. The factor f_{ax} is defined as the ratio of the field at the nose of the profile to the field between parallel plates having the same gap spacing and the same potential difference.

From a study of the profiles for various values of the constants occurring in the expressions a choice was made yielding the values $f_h = 1{\cdot}175$ and $f_{ax} = 1{\cdot}31$. This profile has a reasonably strong section with the ratio of minor to major axes of 0·41. It was compared with a simple circular roll designed to give the same axial distortion factor, both by electrolytic tank measurements and by breakdown test of a section model in compressed gas. Both methods indicated an improvement in radial breakdown strength of 35%, in agreement with the theoretical prediction.

(2.2) The Effect of Eccentricity

In Section 13.2 is given an expression for the field distortion factor f_e on the cylindrical column due to a small eccentricity (point A in Fig. 3). The maximum tolerable eccentricity between the inner column and the intershield or between intershield and tank was taken as $b = \frac{5}{8}$ in, giving $f_e = 1{\cdot}08$. The distortion factor f_e' on the hemispheres due to eccentricity will be between $(1 + b/2r_0)f_e$ and $(f_e)^2$, and a factor 1·15 was assumed.

(2.3) The Intershield Profile

The intershield skirt profile and its field distortion factor were obtained from a conformal transformation which is given in detail in Section 13.3. A choice of constants in the transformation was made which yielded a reasonable profile with a distortion factor $f_s = 1{\cdot}075$ at the point C in Fig. 3.

(2.4) Design of Intershield

With the value of these field distortion factors the optimum dimensions and potential of the intershield can be determined.

If the potentials are V_0 on the terminal spinning, V_1 on the

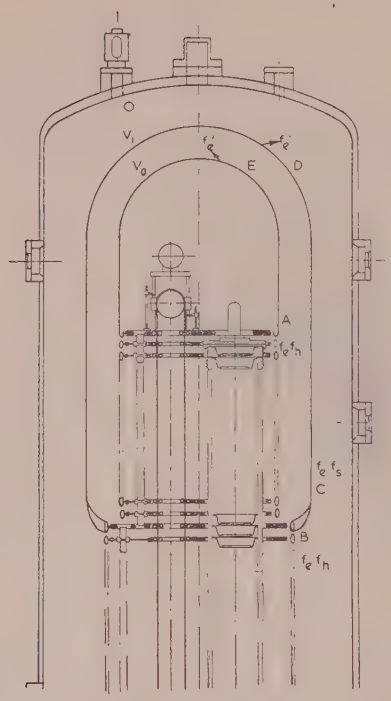


Fig. 3.—A cross-section of the machine, showing fields, potentials and distortion factors.

$E_A = 180$ kV/cm. $E_B = 193$ kV/cm. $f_e = 1{\cdot}075$.
$E_B = 152$ kV/cm. $f_s = 1{\cdot}08$. $V_0 = 5{\cdot}5$ MV.
$E_C = 169$ kV/cm. $f_e' = 1{\cdot}15$. $V_1 = 3{\cdot}24$ MV.
$E_D = 200$ kV/cm. $f_h = 1{\cdot}175$.

intershield and zero on the pressure tank, and if the fields are E_0 and E_1 on the outer surfaces of the terminal and shield respectively, then the elementary theory of concentric spheres readily yields

$$V_0 - V_1 = E_0 r_0 \left(1 - \frac{r_0}{r_1}\right) \quad . \quad . \quad . \quad . \quad (3)$$

$$V_1 = E_1 r_1 \left(1 - \frac{r_1}{R}\right) \quad . \quad . \quad . \quad . \quad (4)$$

The condition is now imposed that E_0 and E_1 must be the same for optimum performance, i.e.

$$E_0 = E_1 = E/f_e' \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where E is the maximum field likely to be encountered in the presence of eccentricity. Hence

$$\frac{V_0 f_e'}{E} = r_0\left(1 - \frac{r_0}{r_1}\right) + r_1\left(1 - \frac{r_1}{R}\right) \quad . \quad . \quad . \quad (6)$$

$$\frac{V_1}{V_0} = \frac{r_1\left(1 - \frac{r_1}{R}\right)}{r_0\left(1 - \frac{r_0}{r_1}\right) + r_1\left(1 - \frac{r_1}{R}\right)} \quad . \quad . \quad . \quad (7)$$

giving the terminal potential for a given field, and the potential division ratio of the stack.

In the cylindrical region, the condition is that the fields at points A and C in Fig. 3 should be the same, i.e.

$$E_0 f_h f_e = E_1 f_s f_e = E \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and the corresponding equations obtained from the theory of concentric cylinders are

$$\frac{V_0 f_e}{E} = \frac{r_0}{f_h}\log_\varepsilon \frac{r_1}{r_0} + \frac{r_1}{f_s}\log_\varepsilon \frac{R}{r_1} \quad . \quad . \quad . \quad (9)$$

$$\frac{V_1}{V_0} = \frac{f_h r_1 \log_\varepsilon \frac{R}{r_1}}{f_s r_0 \log_\varepsilon \frac{r_1}{r_0} + f_h r_1 \log_\varepsilon \frac{R}{r_1}} \quad . \quad . \quad . \quad (10)$$

The ratio R/r_0 has already been chosen as 2. For the intershield radius, eqns. (6) and (9) give a maximum value of V_0/E when $r_0/r_1 = 0{\cdot}682$ for the spherical region and when $r_0/r_1 = 0{\cdot}710$ for the cylindrical region. The value $r_0/r_1 = 0{\cdot}71$ was chosen. The spherical region is slightly away from optimum conditions, but the effect on its value of V_0/E is negligible. The radii being now fixed, the equations give $V_0/ER = 0{\cdot}307$ and $V_1/V_0 = 0{\cdot}590$ for the spherical region, and $V_0/ER = 0{\cdot}348$ and $V_1/V_0 = 0{\cdot}612$ for the cylindrical region. The smaller value of V_0/ER determines the maximum voltage of the generator if E is identified with E_m ($V_0 = 5{\cdot}5$ MV if $E_m = 200$ kV/cm). For the ratio V_1/V_0 the intermediate value 0·60 was chosen. This determines the point at which the intershield joins the stack.

(2.5) The Lower Stack Radius

The curved profile of the intershield reduces the field intensity on the guard-rings immediately below it, and, allowing for the decreasing potential of adjacent rings, the maximum field distortion factor, 0·86, occurs on the third ring below the shield.

The radius r_2 of the lower part of the stack was fixed by the requirement that the field at B in the absence of shielding by the curved profile should be smaller by a safety factor X than the field at A or C. Since the field at B in the absence of shielding is $V_1 f_e f_h / r_2 \log_\varepsilon (r_2/R)$ and the field at C is $V_1 f_e f_s / r_1 \log_\varepsilon (r_1/R)$, the equation for r_2 is thus

$$\frac{r_2}{r_1}\log_\varepsilon \frac{R}{r_2} = \frac{f_h}{X f_s}\log_\varepsilon \frac{R}{r_1} \quad . \quad . \quad . \quad . \quad (11)$$

The shielding effect of the profile might have been sufficient to ensure that breakdown would occur from the intershield rather than from the rings just below it, but as an additional precaution X was given the value 0·87, leading to $r_2/r_1 = 0{\cdot}846$. With the screening due to the profile the field in the worst case of the third ring below the intershield is given by

$$E_B/E_C = 0{\cdot}86X = 0{\cdot}75 \quad . \quad . \quad . \quad . \quad (12)$$

(2.6) The Vertical Field

The stressed length of the stack is 110 in, the largest that can be accommodated in the pressure tank. This gives a mean axial field of 20 kV/cm at the maximum calculated voltage of the generator (5·5 MV). The guard-ring parameters m and a and the plate spacing d were chosen to make the field E_{ax} between rings smaller than the maximum field in the generator by a safety factor of 1·6. The axial field is given by

$$E_{ax} = V f_{ax}/g \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where f_{ax} and g are functions of m and a as already described. These considerations led to a choice of 43 plates with an axial pitch of $2\frac{9}{16}$ in.

Summary of Dimensions

Diameter of tank .. .. $2R = 70{\cdot}5$ in.
Diameter of upper plates .. .. .. $2r_0 = 35{\cdot}25$ in.
Diameter of intershield .. .. .. $2r_1 = 49{\cdot}6$ in.
Diameter of lower plates .. .. .. $2r_2 = 42{\cdot}0$ in.
Ratio of intershield to terminal potential $V_1/V_0 = 0{\cdot}60$.

(3) CONSTRUCTION OF THE GENERATOR

Materials and manufacturing techniques in general follow established engineering principles; where new materials or methods are used, rigorous tests were first carried out to ensure their suitability.

The pressure vessel was designed for 400 lb/in^2. The main cylinder is fastened to the domed base by 48 2¼ in B.S.F. swing bolts, and the gas seal is a tapered spigot and recess with a ¼ in thick rubber gasket.

The belt-driving motors (two of 10 h.p. each), lower pulley assembly and belt-tensioning arrangement are similar to those of the Harwell machine. They are housed in the field-free space between the main base-plate of the machine and the base-plate of the electrostatic column, which is supported 24 in above the main plate by four steel columns.

The generator column consists of metal plates of aluminium–silicon alloy supported on, and insulated from each other by, three insulator pillars. These are built up from a series of insulators screwed end to end, each insulator being 2¼ in long, made from Perspex cemented to brass end-caps using Bakelite vinyl cement. The end-caps were re-machined after cementing to ensure correct length, and when assembled the overall heights of the insulator pillars showed a variation of less than 0·004 in in 9 ft. The change in length per insulator due to cold flow under load has been less than 0·0015 in, but lateral shift up to 0·005 in between the end-caps has occurred in some cases.

Fig. 4 shows the generator plate assembly. Each plate carries

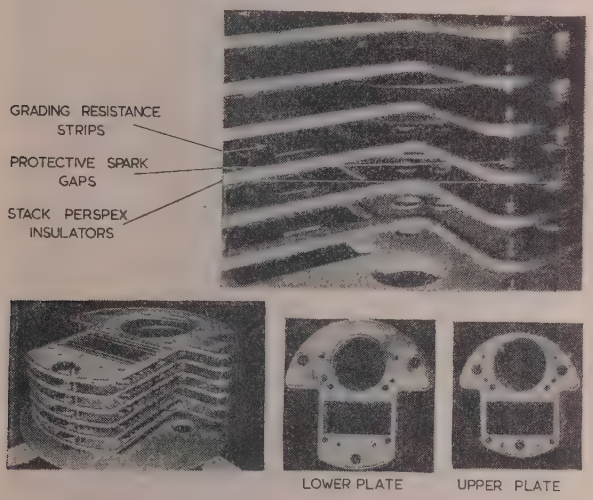


Fig. 4.—The generator column plate assembly.

four belt guides, each made of ¾ in diameter Perspex rod into the face of which a glass spacer rod of ⅜ in diameter is cemented. The plates were designed with "cut-away" sides to facilitate removal of the resistor units and the inner belt guides.

The guard-rings are made from stainless-steel tubing drawn to the required profile and then rolled into hoops of 35 in and 42 in diameter. Each ring is keyed to its plate by three long keys welded to the inside of the ring engaging with keyways machined in the periphery of the plate. The electrostatic column with guard-rings fitted is shown in Fig. 5.

The column diameter is changed from 42 in to 35 in by a special "interchange" plate which supports the intermediate shield and allows the pitch circle of the insulator pillars to be changed. The top plate of the stack carries the top terminal shield, the top belt pulley, and all the equipment associated with the ion source.

The intershield is 80 in long and 0·083 in thick, and the terminal shield is 40 in long and 0·060 in thick. Both are made from stainless-steel sheet, and each is welded at the lower end to a guard-ring with the standard key fitting to the plates.

The resistance chain from top to bottom of the stack is mounted between the vertical runs of the belt, as shown in Fig. 4. It consists of 44 separate units connecting the column plates together, each unit consisting of 37 20-megohm composition resistors mounted between two Perspex sheets. The resistors are of 2-watt rating and are normally under-run, but they suffer damage from time to time under transient over-voltages.

(3.1) Gas System

The gas normally used is dry nitrogen with 10% by volume of Freon (CCl_2F_2) at a pressure of about 150 lb/in^2. When not in use the gas is stored in reservoirs equal in volume to the generator tank, thus avoiding the possibility of condensation of the Freon during pumping.

The two-stage compressor has a pumping speed of 10 ft^3/min corresponding to a half-value period of 50 min when emptying the tank. An SR5 vacuum pump, with five times this speed, is used to evacuate and dry the machine before filling and to remove the last atmosphere of gas before opening the machine.

The gas is continually circulated through a water-jacket cooler and two large silica-gel driers external to the tank. The circulation is done by a Rootes-type impeller mounted with its motor on the main base-plate, the hot gas being drawn from the top terminal through a system of Perspex tubes fitted between the stack plates. The impeller speed is estimated at 40 ft^3/min, which suffices to keep the terminal equipment within 30° C of the ambient temperature. With 2 kW dissipated in the terminal equipment and a pressure of 175 lb/in^2, the equilibrium temperature at the base of the machine is about 40° C. The frost point of the gas is maintained at about −50° C.

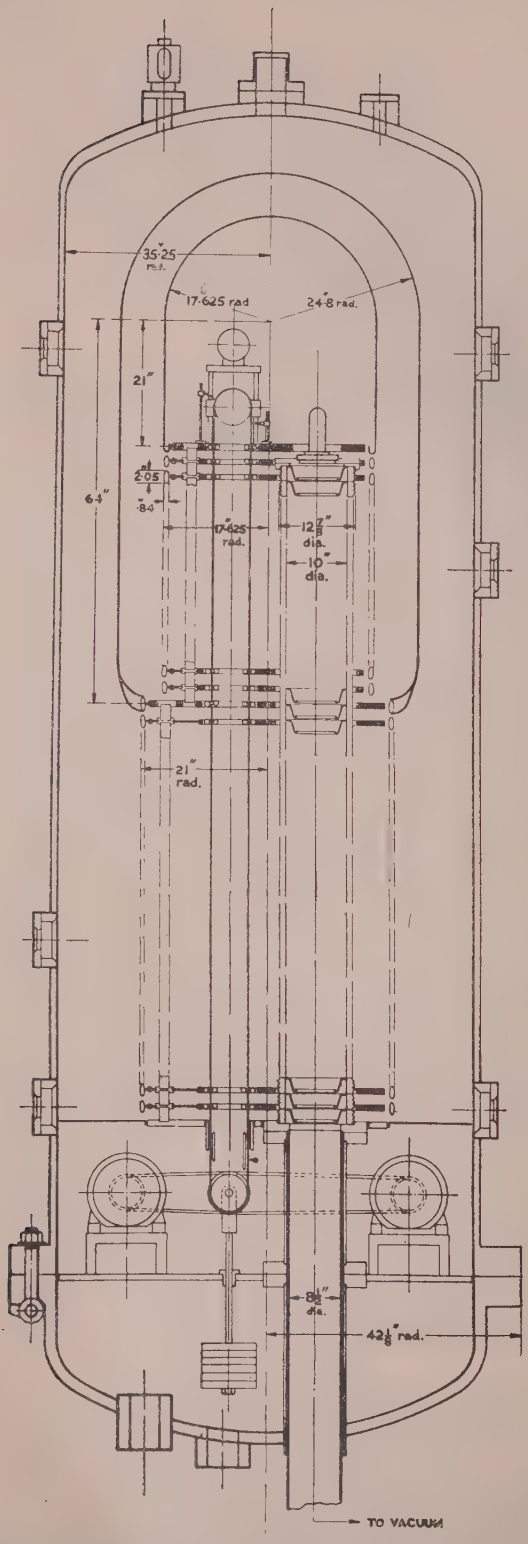
(3.2) Control System

Three control stations are provided. Two local stations are on the generator floor and in the target room and are used for testing and setting up experiments. Neither can be used if the radiation intensity in the target room exceeds 0·025 r/h or when the integrated dose has exceeded a value of 0·1 r during the day. The change-over to the remote control room can be effected only if the safety interlock sequence is complete. An important feature is that the change of control position does not affect any of the control settings.

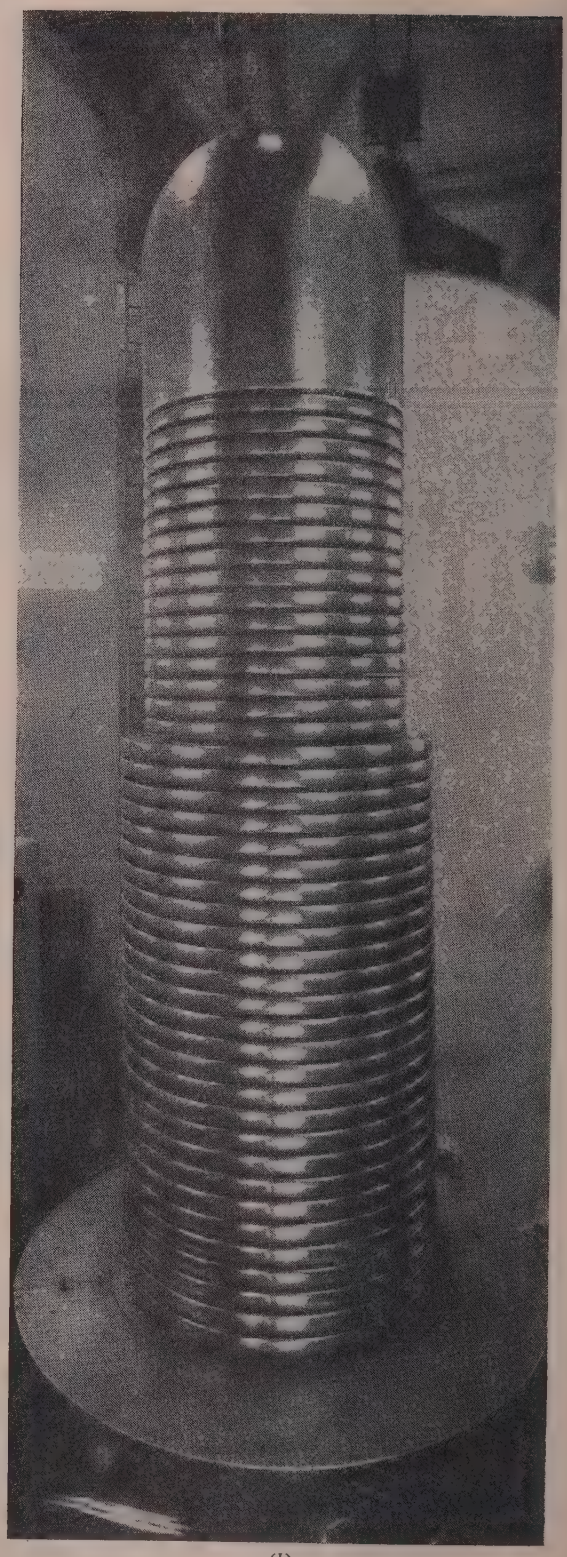
The control sequences are automatic where possible, and, in particular, the vacuum system is fully automatic and will re-cycle itself following a power or water failure. Audible warning precedes the starting of the belt motors, and spray current cannot be applied to a stationary belt. The usual safety interlocks are provided on all high-voltage equipment, and an emergency electromagnetic braking system enables the belt to be stopped in 4 sec.

The voltage control of the machine is described in greater detail in Section 8, but from the operational point of view the generator can be brought to a given voltage merely by the operation of a set of pushbuttons. The remote controls for the ion source and other top terminal equipment are also on the main control desk.

In addition to the actual operational controls, it is possible from the control positions to monitor continuously the various voltages and currents in the top terminal by means of the telemetering equipment, to measure the spray, stack, beam and



(*a*)



(*b*)

Fig. 5.—(*a*) Cross-section of the machine. (*b*) The electrostatic column with guard-rings and terminal shield.

target currents, to measure the generator voltage with the generating voltmeter and check the operation of the vacuum system and the voltage stabilizing system.

All the wiring, both inside the bell and to the control panels, is carefully screened, and all instruments are protected against the violent transients which appear when the generator sparks over.

(4) GENERATOR PERFORMANCE

Ease of assembly and general convenience of design have been confirmed in practice. The machine has been completely dismantled several times, which takes one day after the pressure tank has been removed. Several days are required for rebuilding and testing.

The power consumption of the machine should be kept to a minimum to avoid undue heating of the gas. By removing the cooling fans from the main motors, a saving of 1·5 kW at 150 lb/in^2 and 3·7 kW at 400 lb/in^2 was effected, the temperature rise of the yokes then being 17° C. Table 1 shows the power absorbed in different parts of the machine at the pressure mentioned.

Table 1

POWER ABSORBED BY DIFFERENT PARTS OF THE MACHINE

Part	Power	
	at 150 lb/in^2	at 400 lb/in^2
	kW	kW
Motors (without fans)	2·1	4·0
Belt and pulleys	1·5	1·5
Terminal equipment	1·2	1·2
Gas circulator and motor	0·8	2·1
Total	5·6	8·8

The generator was first tested without an ion accelerator tube, and the terminal voltage was initially estimated by measuring the current flowing at the earthed end of the resistance chain. The voltage so calculated should be treated only as approximate for three reasons. First, the value of the potential-dividing resistor has been shown to vary with time; eight measurements have shown that after an initial drop of 9% the most probable value over the last two years is $(31{\cdot}3 \pm 0{\cdot}5) \times 10^9$ ohms. Secondly, the resistance value has been shown to depend slightly on the value of the current flowing. Lastly, as will be discussed in Section 7, careful measurements indicate that the current in the resistor chain is non-uniform to the extent of ±7 or 8% even in the absence of the intershield, and with the intershield present the current flowing in the upper part of the resistor chain can be markedly different from the metered current at earth potential owing to ionization effects in the dielectric gas. However, later measurements indicate that, provided the generator is operated so that the spray current is never greater than 130% of the metered resistor current, the method can be used to estimate the terminal potential to an accuracy of ±4%.

(4.1) Current Performance

Resistor current and the quantity (spray current − resistor current) = I_Δ were plotted against generator voltage for different gas pressures. These characteristics were similar for all pressures, and the curves for 150 lb/in^2 are shown in Fig. 6 for four different types of guide bars. It is seen that with Perspex bars the equivalent stack resistance remains sensibly constant up to voltages of about 4 MV, whereas the metal guide bars cause a 21% decrease in effective resistance at 4 MV. The known non-linearity of the resistor material accounts for about one-third of the decrease in equivalent stack resistance. It is suggested that the remaining non-linearity of the resistor current curves is due to leakage of current by corona from the belt to the guide bars and hence to the metal plates. Such current would be measured as resistor current and would be least for Perspex guides and greatest for metal ones. Later measurements suggest that under most conditions current sprayed on to the belt is efficiently delivered to the terminal. Hence the quantity I_Δ is a good estimate of current passing radially through the gas, and it is concluded that radial corona currents of about 10 μA at 3 MV and about 45 μA at 4·4 MV existed under these conditions. Curves D of Fig. 6 were taken with a vacuum tube, and the marked increase in the quantity I_Δ in this case is due to photo-ionization produced in the gas by X-rays generated in the vacuum tube. Fig. 7 also shows that the radial corona losses increase rapidly for resistor currents above 120 μA (about 3·6 MV), and that the loss is least for the highest pressure.

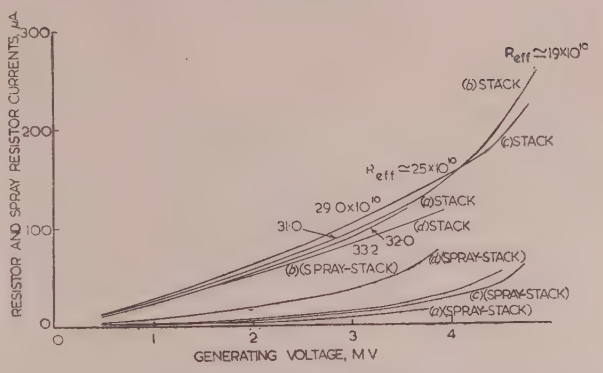


Fig. 6.—Variation of resistor current and (spray-resistor) current for various belt guide materials.

Pressure, 150 lb/in^2.

A. Zircon–Araldite shielded.
B. Metal.
C. Alternate metal and insulating.
D. Perspex.

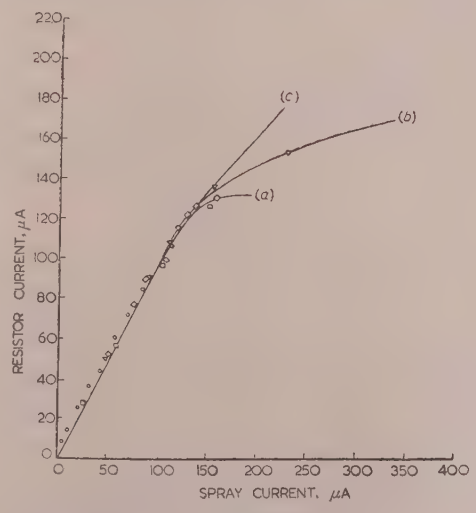


Fig. 7.—Relation between resistor current and spray current at various pressures.

(*a*) 75 lb/in^2. (*b*) 100 lb/in^2. (*c*) 175 lb/in^2.

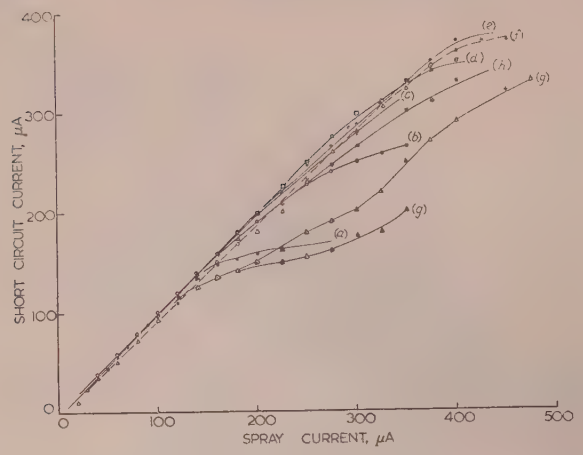


Fig. 8.—Relation between short-circuit current and spray current at various pressures.

Alternate insulated and metal guide bars.

(*a*) 15 lb/in². (*b*) 30 lb/in². (*c*) 60 lb/in². (*d*) 90 lb/in².
(*e*) 120 lb/in². (*f*) 150 lb/in². (*g*) 180 lb/in². (*h*) 135 lb/in².

Fig. 8 shows the short-circuit characteristics of the generator, and it is seen that, for pressures between 60 and 150 lb/in², currents up to 300 μA are delivered by the belt to the terminal with little or no loss in the transfer. Above 300 μA spray current the transfer becomes less efficient owing to losses at the needles, and from the charged-belt surface to the metal plates and between 300 and 400 μA spray current only about 90% can be obtained as short-circuit current.

The transfer efficiency is pressure-dependent, falling steeply for pressures in excess of 150 lb/in² as shown by the curve for 180 lb/in.² The pressure dependence is due to the fact that the discharge at the needles changes from corona to constricted-spark type of discharge when the gas pressure exceeds about 150 lb/in.² This leads to a markedly non-uniform distribution of charge on the belt resulting in excessive electric fields and surface flashover as well as sparking through the belt to the pulley or back plate.

Since there is an advantage in increasing the pressure in the main gaps of the generator above 150 lb/in² to reduce radial corona losses, the spray combs have been redesigned in an attempt to control the local field pattern so that a corona discharge can be obtained with gas pressures of up to 200 lb/in².

(4.2) Voltage Performance

A generating voltmeter of the rotating-vane type was mounted in the top of the pressure tank and served to measure the voltage of the intershield when this was present or the terminal voltage in its absence. There are two sets of eight interlaced plates mounted behind the rotating vane, and the instrument produces a 400 c/s signal voltage. Detecting and calibrating circuits are similar to those used by Fortescue and Hall,[1] giving a rectified signal output independent of the vane speed. Their method of altering the sensitivity for calibration purposes was used.

An improvement in the amplifier attached to the generating voltmeter has been incorporated by Churchill and Millar.[3] By separately amplifying the antiphase (sine-wave) component and the in-phase (transient) component, the response time of the instrument has been reduced from ½ sec to 10 microsec.

The voltage limit of the generator was determined as a function of gas pressure using the generating voltmeter to measure the intershield potential. The terminal voltage was then estimated by multiplying the intershield potential by the factor R_t/R_i, where R_t is the value of the potential-dividing resistance between the terminal and earth, and R_i is the value of that part of the resistance between intershield and earth. The measured value, 1·54, of this ratio remained constant over the current range. The calibration curve was cross-checked, after the accelerator tube had been installed, by detecting the γ resonance in the $Be^9(p, \gamma)B^{10}$ reaction, which occurs at a proton energy of 2·56 MeV[4]. The calibration gave an energy of 2·62 MeV, and an error of 2·3%.

In some cases estimates of the terminal voltage were made from the resistor current using a measured value of 32 000 megohms and allowing for the non-linearity of the resistor. It was found that these estimates agreed with the generating-voltmeter estimates within ±4% for positive polarity up to 3 MV and ±10% for negative polarity up to 4 MV.

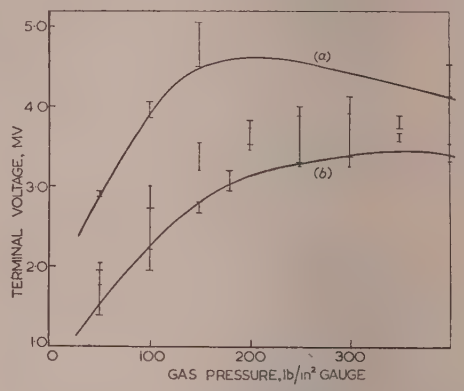


Fig. 9.—Terminal voltage as a function of gas pressure for negative polarity, with intershield.

Temperature, 40° C.
(*a*) Nitrogen with 10% Freon. (*b*) Nitrogen.

Fig. 9 shows the results of several runs with negative polarity with both nitrogen and nitrogen-Freon mixture; the generator limit of 5·5 MV for the nitrogen-Freon mixture was set by sparking down the belt.

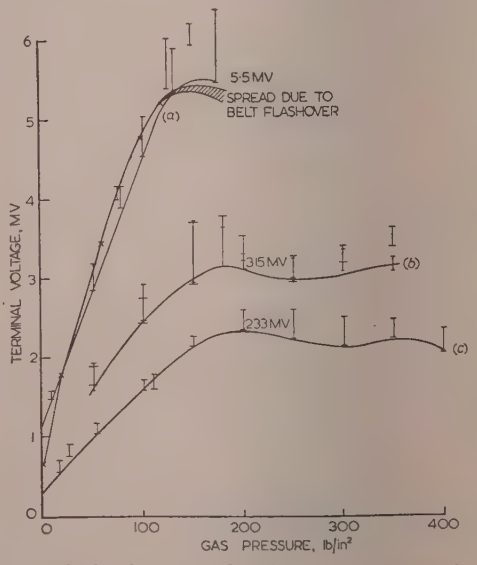


Fig. 10.—Terminal voltage as a function of gas pressure for positive polarity.

(*a*) Nitrogen with 10% Freon. (*b*) Nitrogen.
(*c*) Nitrogen with no intershield. Temperature, 40°C.

Fig. 10 shows the performance obtained with positive polarity for the three conditions of interest.

The maximum voltage without the intershield is 2·3 MV in clean nitrogen, and the improvement due to the intershield is about 30% compared with the theoretically expected improvement of 36%.

The addition of Freon increases the maximum voltage with intershield by approximately 70% from about 3·2 MV to about 5·5 MV.

There is no sensible improvement of maximum voltage as the gas pressure is increased from 200 to 400 lb/in^2. It can be concluded from the current curves that corona losses across the gas are not in excess of 20 μA for terminal voltages up to 4 MV, and hence it appears from the results of Section 2 that the gas will withstand peak electric fields of approximately 160 kV/cm over the main gaps, and the maximum fields before breakdown are approximately 210 kV/cm.

(5) THE ACCELERATOR TUBE

The accelerator tube, described by Chick and Miranda,[5] is made up of 43 porcelain rings of 10 in internal diameter separated by steel diaphragms to which the electrodes are clamped, the whole being sealed together with cold-setting Araldite cement. The pumping speed of the tube computed roughly, regarding it as a series of tubes 2·56 in long and of 10 in diameter separated by discs with 7 in apertures, is 110 litres/sec for air. This is reduced to 80 litres/sec by the impedance of the pumping manifold.

The first design had aluminium electrodes dished as in Fig. 11

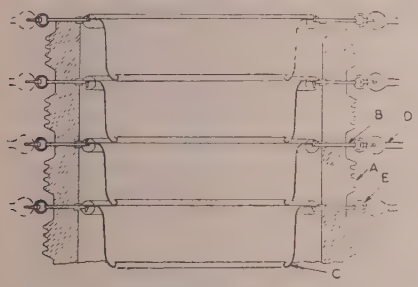


Fig. 11.—The first tube design (aluminium electrodes).

A. Porcelain rings.
B. Steel diaphragms.
C. Accelerating electrodes.
D. Stack plates.
E. Conducting rubber rings.

to protect the tube wall from direct exposure to the beam. The maximum field strengths at both anode and cathode of each electrode were determined from field plotting, and found to be 1·65V volts/cm, where V, the voltage per gap, is 95 kV when the generator voltage is 4 MV.

Each electrode was metallically connected to the corresponding generator plate, to grade the potential uniformly along the tube. The two topmost electrodes were used for focusing the beam and were connected together to a 0–40 kV power supply in the top terminal. Ray tracing and subsequent experience showed that by this means it was possible to produce a 3 MeV beam with a diameter less than 1 mm at a target 18 ft below the source.

After a short conditioning time this tube withstood 3·4 MV, but after 6 months internal flashover of the tube occurred at 2·3 MV. When the tube was dismantled the following causes of deterioration were found:

(*a*) Heavy damaging discharges between electrodes had volatilized aluminium from the anode surfaces and spread it over the cathode and porcelain surfaces.

(*b*) Surface flashover along the porcelain had burnt the excess sealing resin, which was extruded around the inside of the joint, spreading carbon over the porcelain, which led to further tracking.

Investigation of these phenomena led to the design of a second accelerator tube. It was shown that the spark-gaps provided between the generator plates to protect the tube against overvoltage surges failed to act because the time-lag of the breakdown of a gap in high-pressure gas is much greater than that in a vacuum. A study was made[6] of the equivalent circuit of the generator as a capacitance network, in the hope of finding the cause of these surges. If a radial breakdown occurs between intershield and tank, the potential on the adjacent gaps changes from V to $-3{\cdot}6\,V$, corresponding to -350 kV if operating at 4 MV. This effect is due to the small capacitance between plates and tank and is missed by any simplified calculation which ignores these capacitances. If one gap breaks down, the neighbouring gap potential changes to $1{\cdot}09V$; if 10 adjacent gaps break down, the next gap potential changes to $1{\cdot}85V$, much more than the steady gap potential.

In the second tube, which is shown in Fig. 12, the electrodes were made of stainless steel, a metal which is not appreciably

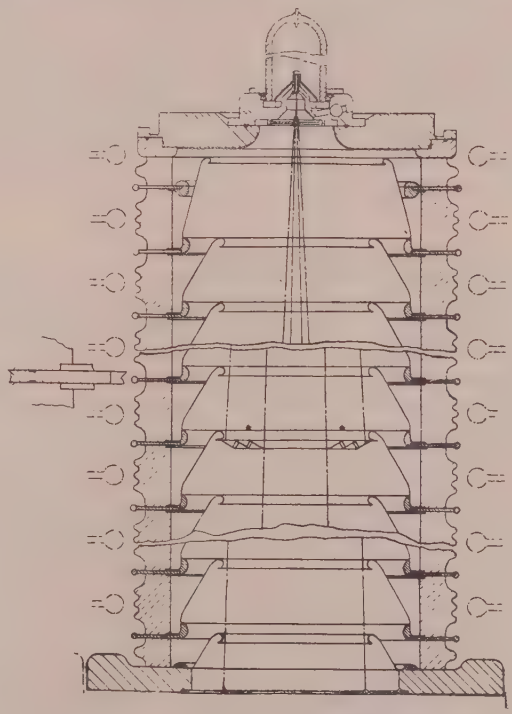


Fig. 12.—The second tube design (steel electrodes).

damaged by the heavy vacuum discharges, and the design has been inverted to protect the porcelains from any metal volatilized from the anode surface. The inter-electrode gap has been made larger, and the maximum field is estimated at 1·15V volts/cm on the anode and 0·95V volts/cm on the cathode. Precautions have been taken to prevent excess Araldite from entering the tube. The diaphragms were connected to the generator through a high resistance, which, with the inter-electrode capacitance, gives sufficient time-lag to enable the protective spark-gaps to function in the event of surges.

This tube has now been operating for 1 000 hours and shows no sign of deterioration in performance. On a recent occasion when the tube was opened for leak repairing, it was observed that, although there were signs of a very few inter-electrode

sparks, there was no damage to electrode surfaces or porcelain walls. The maximum voltage reached after careful conditioning was 3·8 MV.

(5.1) Vacuum Equipment

The vacuum system external to the accelerator tube comprises the target manifold and vacuum ducting leading to two 8 in oil diffusion pumps in parallel, each with a speed of 500 litres/sec at 10^{-5} mm Hg. These are backed by a single-stage SR5 rotary pump rated at 13 litres/sec at 10^{-1} mm Hg. To reduce back-streaming from the diffusion pumps, their heater power is reduced from 1 800 to 1 500 watts and their baffles are refrigerated to $-40°$ C.

The pumped volume is 900 litres, and a workable vacuum is obtained 20 hours after starting the backing pump. The ultimate vacuum pressure as measured by an ionization gauge in the target manifold is between 7×10^{-7} and 2×10^{-6} mm Hg (depending on temperature), rising to between 3 and 4×10^{-6} mm Hg when the ion source is running. The system has a background leak of 0·04 litre micron/sec. No leak in excess of 0·1 litre micron/sec at atmospheric pressure is permitted in the part of the system exposed to high-pressure gas, as such leaks tend to increase very rapidly with pressure.

The vacuum plant works continuously and unattended, and is protected against failure in electric power or water supply and against development of serious vacuum leaks. The switching on of the pumps and operation of pneumatic flap valves in both the high-vacuum and backing sides are completely automatic and occur in correct sequence whether the tube is initially at atmospheric pressure or under vacuum.

(6) TERMINAL EQUIPMENT

Within the terminal shield is installed the necessary equipment for removing the charge from the generator belt, for producing protons or deuterons as required, and for focusing the ion beam over the energy range of the generator.

Electric power is provided in the terminal by a 2000 c/s alternator driven from the main charging belt and excited by a 24-volt d.c. generator. The a.c. line voltage is stabilized to $\pm 2\%$. Three variable d.c. supplies are provided from the alternator: 0–2 kV, 250 mA for the r.f. oscillator anode supply; 0–5 kV, 10 mA for the ion source extracting potential; and a negative supply 0–40 kV at 0·5 mA used to control the potential of the first tube electrode as a means of focusing the ion beam. The voltage levels of these supplies are controlled by two Variacs, each with two sliders, driven by reversible d.c. motors, while a motor-driven wafer switch allows a choice of ions.

High-pressure storage cylinders are provided for hydrogen and deuterium, the flow of these to the ion source being controlled by palladium leaks directly heated by current controlled by a motor-driven Variac.

The ion source is of the r.f. discharge type first described by Herzberg[7] in 1927 and developed by Thonemann[8] and Bayley and Ward.[9]

When an ion beam is being produced the envelope contains hydrogen or deuterium at a pressure of approximately 2×10^{-2} mm Hg. This pressure is metered at the control desk, a feature which has proved valuable in obtaining repeatable ion source conditions.

From the ion source envelope a Duralumin canal 0·062 in in diameter and 0·75 in long leads to the main accelerator tube. The canal has a gas consumption of 10 cm^3/h at standard temperature and pressure. The ion source oscillator employs a balanced Hartley circuit with grid current biasing to provide the 20 Mc/s output. Since the frequency is not critical the tank coil can be directly coupled to the ion source, some 200 watts being absorbed by the gas load. The two oscillator valves, electrically similar to the R.C.A. type 813, were specially made to withstand the high external gas pressure and will handle up to 500 watts input.

The power supplies in the high-voltage terminal are controllable from outside the generator, and remote indication of various currents, voltages and gas pressures inside the high-voltage electrode are provided.

Great flexibility was obtained by using modulated light beams as links between the high-voltage electrode and the base of the generator, both for the up-going control system and for the down-coming metering system. B. Millar[10] has described the electronic control and metering system in detail.

In the remote-control system, the light from a crater-lamp light source can be modulated by any one of seven audio frequencies by means of seven keys on the control desk. Pressing a key in the reverse direction modulates the light with an additional "reversing" frequency. In the terminal electrode the light is received on an electron-multiplier photocell. The a.f. output from this is identified by a set of channel-discriminating tuned circuits, and after suitable amplification and rectification operates the appropriate one of a set of relays. In the majority of the control channels the relays energize 24-V d.c. motors which are arranged to drive Variacs or move switches.

To provide an indication of the control settings, a telemetering system was developed largely by Millar,[10] which uses a pulse-modulated light beam to give an apparently simultaneous and continuous indication at the control desk of the values of five variables in the high-voltage electrode.

The flexibility of the system has proved to be a great advantage, and, by diverting two relatively unimportant control channels to switch the telemeter input, it has been possible to meter up to 17 quantities in the terminal, making possible the detailed current balances discussed in Section 7.

The five required electrical quantities are transformed into time-modulated pulses by using ten special saturable reactors or peaking transformers. Five of the reactors produce "marker" pulses, which occur at equal intervals of 2·64 millisec and define the scale zero points. The other five reactors produce the required "pointer" pulses, and each of these is delayed from its associated marker pulse by a time proportional to the current to be measured. The information is sampled at a repetition rate of about 60 times a second. The reactor output pulses, which are about 30 microsec wide and of 20 mV amplitude, are amplified and used to modulate a cathode-ray-tube light source.

The calibration of the system depends mainly on the rate of rise of a sawtooth current waveform and is accurate to better than 2% over long periods.

The receiver displays the pulse train on a 5-line raster on cathode-ray tubes at the control desks. Each line of the raster is initiated by a channel marker pulse, and the information is visible as a vertical pulse displaced along the line.

The quantities measured include the oscillator anode potential and current, the extractor electrode potential and current, the gas pressure in the source, the focus electrode potential, the current stripped from the belt, the current flowing into the resistor chain, the current flowing to the anode of the tube and any current leaving the terminal through the insulating gas. Some of these currents are too small to give full modulation of the saturable reactors, and cathode-follower amplifiers are then used. In these cases the accuracy is limited to about 5% by the amplifier input resistors, which are of the carbon type.

(7) ACCELERATOR PERFORMANCE

As the generator voltage is increased the tube becomes momentarily conducting above a certain threshold voltage, but

recovers after the passage of a small charge, presumably because of the slight drop in generator voltage. These pulses of conduction are repeated at regular intervals. Above this threshold additional belt current produces relatively little rise in voltage but increases the frequency of these pulses. By prolonged running in the pulse conduction state the threshold rises, the conduction at a given voltage diminishes, and the voltage can be taken higher.

Since the pulse type of breakdown does not short-circuit the generator voltage completely, the tube can be used in this state, but with a limited attainable voltage for two main reasons. First, the conduction consists of electrons and negative ions travelling up the tube and positive ions travelling down. These strike the tube electrodes so as to cause an uneven distribution of potential along the tube. Secondly, electrons striking the top plate of the tube produce X-rays which ionize the high-pressure gas, causing an ionization current from top terminal to tank. This current upsets the uniformity of the potential distribution because the current from top terminal to intershield is usually not the same as that from intershield to tank. The tank has been insulated from earth and the ionization current measured. This current may exceed 100 μA under extreme conditions with a proton beam and an electron beam both present, and 250 μA has been observed with the intershield removed. This and the tube current constitute an unnecessary load on the belt and are thus apt to limit machine performance. When the proton beam and electron beam are switched off it is observed that, when the threshold voltage is reached, pulse conduction, X-rays and tank current all appear simultaneously.

An interesting theory of the mechanism of high-voltage breakdown in a vacuum has been put forward by Harris of Queen Mary College,[11] who has studied pulse conduction in a short tube at voltages up to 800 kV. Positive ions, after acceleration down the tube, strike the bottom and eject neutral particles, ions and electrons. Some neutral particles, emitted backwards, reach the top, and Harris suggests that these are mainly responsible for ejecting further positive ions. If the secondary emission coefficients for the production of positive ions by neutral atoms and neutral atoms by positive ions is greater than unity, the process is self-maintaining. On this theory (unlike previous particle-exchange theories) the electrons and negative ions are merely by-products of the chain reaction.

In order to break the chain reaction causing breakdown, a positive-ion suppressing field was produced over the surface of the top plate of the tube by connecting the focusing electrode, and the one below it, to a 0–40 kV supply, positive with respect to the top plate. Applying this retarding field increased the breakdown voltage by an amount ranging between 5 and 15% depending on the degree of conditioning of the tube, the vacuum pressure, and whether the proton beam was on or off. With no beam and a good vacuum, the suppressor would generally allow the voltage to be raised up to the ultimate breakdown limit (where the generator voltage falls catastrophically to zero) without any tube current or X-ray production.

Another effect noticed by other workers and confirmed with the Aldermaston generator is that the breakdown voltage is improved by raising the pressure in the tube from 2×10^{-6} to about 3×10^{-5} mm Hg. This is not so effective as the suppressing field. The two effects generally add if used together. The use of such a high tube pressure is not a good solution, however, as it causes considerable scattering of the main proton beam.

Since the existence of tube conduction and X-rays profoundly modifies the currents in different parts of the generator, a complete investigation of currents was made under various operating conditions, making use of the telemeter facilities for measuring

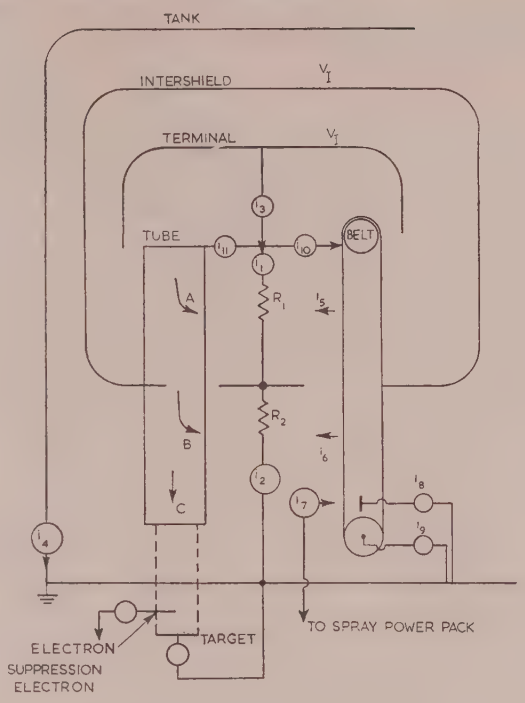


Fig. 13.—Schematic of the machine.

as many currents in the top terminal as possible. Fig. 13 shows a schematic of the machine. There are three ways in which the measured currents may balance; lack of balance gives information on current lost to the generator plates or the tube walls, both of which cannot be measured directly. First, the current sprayed on to the belt may balance that taken off at the top, plus that lost to the lower pulley and charging plate, all of which are measured. Any loss is attributable to charge leaving the belt and reaching the generator plates as it travels upwards. With the intershield in position the loss is negligible, but with the intershield removed up to 30% of the spray current may be lost to the plates. Secondly, the current removed from the belt at the top should flow partly into the resistor chain, partly into the tube, and partly to the terminal spinning and across the gas to the intershield. These are measured separately, and the balance constitutes a check on the accuracy of the telemeter. Thirdly, the spray current should balance all currents returning to earth at the base of the machine; these are the currents to the lower pulley, charging plate and tank, the current from the resistor chain, and the tube current after subtraction of loss to the tube wall (since this will be accounted for in the resistor current). As all these currents are measured, except the loss to the walls, this can be readily calculated.

As a result of these measurements reported in detail by Trott and others[12] it has been possible to postulate an equivalent-circuit model for the generator, the main features of which indicate that the two corona components of the gas current pass between the main terminal electrodes and not from the electrostatic column; and, for any potential, corona current across the outer gap is about six times the corona current across the inner gap.

About 90% of the ionization currents in each gap come from the terminal shields, and the ratio of ionization currents in the outer and inner gaps is about 3.

The current loss per unit area to the tube walls appears to be

constant down the length of the tube, and with no ion beam this loss was usually negative, increasing in magnitude with potential. With an ion beam the loss was positive, i.e. like a positive current to the wall, and increased in magnitude with the beam current.

If there is no tube current (i.e. no ion or electron beam and no pulse conduction) and no corona current across the gas, the measured current entering the top of the resistor chain is equal to that leaving the bottom. Under these conditions the *IR* product for the lower half of the stack gives the intershield potential in agreement with the generating voltmeter reading, and the *IR* product for the whole of the stack gives the top terminal potential reliably. At the highest voltages, however, even when the tube is not conducting, there is usually corona current across the gas, and it is observed that the difference between the current arriving at the tank and that leaving the terminal spinning just balances the difference between the upper and lower resistor currents. The terminal potential is then given by the sum of the separate *IR* products for the upper and lower parts of the stack. Under such conditions the peak performance was found to be 3·8 MV.

When there is considerable tube current, further small corrections are applied for current loss to the walls of the tube (calculated as above) and for ionization current flowing from the stack guard-rings. This current is assumed to be about one-tenth of the total ionization current across the gas, which is determined by measuring the currents from the terminal and to the tank under working conditions, and subtracting the values obtained at the same voltage when the tube is not conducting (i.e. the corona components). In this way the peak performance under beam conditions was found to be also 3·8 MV.

As there is slight uncertainty about the absolute value of the resistance under working conditions, more reliable results are obtained by using the current analysis to calculate the voltage division ratio of the stack, and applying this ratio to the observed generating voltmeter reading.

The validity of this method of calculating terminal potential has been proved by taking current measurements while the accelerator is being held on the $Be^9(p, \gamma)$ resonance at 2·57 MV, and more recently by actual measurement of the beam energy with the electrostatic analyser.

The relationship between ionization current through the gas and electron current up the tube was examined at a fixed voltage (2 MV) by measuring the tank current for various values of current from the electron gun mounted at the bottom of the tube in the absence of protons and pulse conduction. It was found that the tank current was proportional to the electron current, and the ratio (microamperes of tank current per microampere of electrons) is proportional to the gas pressure, ranging from 1·0 at 70 lb/in² to 3·0 at 200 lb/in² (absolute).

A pair of ionization chambers well collimated with lead shields were used to detect horizontally propagated X-radiation at various vertical distances from the tube anode. These chambers were placed at the nearest point to the tube just outside the tank wall. The distribution of X-ray intensity with the generator running at 2·7 MV with no ion beam and no control electron beam is shown in Fig. 14. It indicates that the X-rays come predominantly from the top of the tube. A check chamber mounted in the atmosphere directly above the tube indicated an intensity of 2·5 r/h during these exposures while the target room intensity was measured as 0·01 r/h. Several absorption measurements were carried out at a fixed generator voltage to compare the energy of the X-rays produced by the tube with that produced by the control electron beam. The results indicated that the tube X-rays had an effective energy of 1·24 ± 0·3 MeV whereas for the X-rays due to the electron beam an energy of 1·55 ± 0·18 MeV was estimated. The difference is not significant, and it is concluded that in general the tube X-rays are produced by electrons which travel the whole length of the tube.

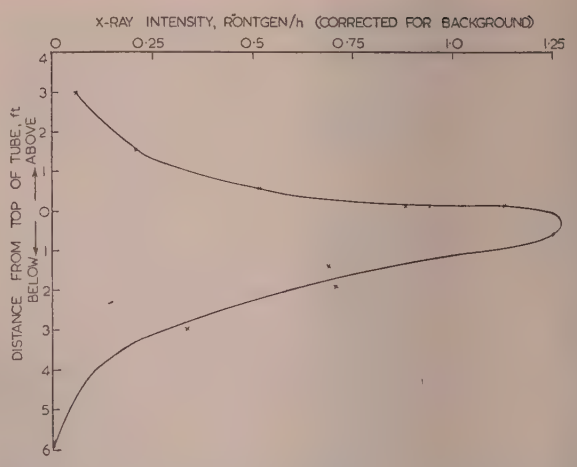


Fig. 14.—X-ray intensity in radial direction as a function of height from the top of the tube; no beam and no electron gun current.

(8) VOLTAGE STABILITY

An electrostatic beam-analysing system was chosen in preference to a magnetic analyser, since it was felt that an electrostatic system would more conveniently allow of absolute measurement of beam energy and would be more flexible as an element in the voltage-stabilizing servo system.

Beams up to 3·25 MeV have been passed through the analyser, and currents of over 20 μA have been recorded on a target placed immediately after the exit slit using an energy resolution of 1 200.

The main control for setting and measuring the analyser electrode potentials is a precision potentiometer which can be set remotely. The voltage across this potentiometer is controlled by the voltage of a bank of eight standard cells by a suitable servo loop, as described by Bailey.[13]

A separate servo system compares a small fraction of the potential between the analyser electrodes with the potential set by the slider of the precision potentiometer, and uses the error signal to correct the electrode potential by altering the impedance of series valves for small changes and by means of a motor-driven Variac for large changes.

Great care is taken to determine the ratio of the high-voltage dividers, and the resulting uncertainty in the measurement of the deflector potentials is estimated to be rather less than 5 parts in 10^4.

The analyser geometry and its electrode potential define with respect to the standard cells a required beam energy which can be selected by setting the precision potentiometer. Any deviation of the beam energy from the required value will result in changes of the beam currents falling on the exit slit jaws.

The main servo loop is completed by suitably amplifying an error signal derived from the jaw currents and using it to modulate a control electron beam which passes up the ion accelerator tube, as described by Chick and Inall.[14] The mean value of the control electron beam is 25 μA, which at 3·2 MeV creates an X-ray field in the target room of 0·006 r/h.

The three time-constants in this loop are such that a maximum gain of about 5×10^4 is permissible before instability occurs. Over long times, the unstabilized generator voltage varies by

less than $\pm 10\%$, thus a loop gain of about 1 000 is required and the system is stable; the jaw amplifier has an effective time-constant of 12 microsec and the r.f. link to the electron-gun grid has a time-constant of 8 microsec; thus it is seen that the loop is capable of correcting fast disturbances.

The range of energy deviation controlled by the analyser loop is limited to about $\pm 1\%$ of the beam energy, because with greater deviations the beam would strike the electrodes and no correcting signal would be available.

In order to deal with large changes a second servo loop is incorporated in which the signal from the generating voltmeter is compared with a reference voltage proportional to the precision potentiometer setting. The difference signal is suitably amplified and used to modulate the control electron beam if, and only if, the beam energy deviation is greater than 1%. When the analyser is not in use, however, the circuits allow the 1% "backlash" to be reduced so that the generator can be stabilized to one part in 1 000 from the generating voltmeter.

The second loop has a time-constant of 8 sec associated with the generator and three others in the range 10–30 microsec, thus the loop is stable for a gain much greater than the required value of 10–100.

Since the energy range over which the electron beam can control the generator is limited, it is necessary to adjust the spray current manually to cover the full energy range. For any selected beam current and energy, the spray current is set manually so that the mean value of the electron current is approximately correct. The spray current is initially stabilized at the value set by the manual control, but this value is modified through a subsidiary loop by slow changes in the electron beam current. Thus, if a major flashover causes the electron beam current to be removed for several seconds, the subsidiary loop will cause the maximum value of charging current compatible with the manual setting to be applied to the belt.

An overall check on the machine stability is given by noting that with the analyser set for a resolution of 1 200, the amplitude of the target current varies by less than 5% over periods of 30 min.

(9) BEAM ENERGY MEASUREMENT

The analyser, described in detail by Hunt,[15] has a mean radius of $36 \cdot 0000 \pm 0 \cdot 0005$ in and an electrode separation of $0 \cdot 3752 \pm 0 \cdot 0003$ in. The deflecting angle is $63 \cdot 2°$ in order to focus a parallel incident beam on to a 0·015 in wide exit slit. The deflecting potentials required are about +40 kV and −40 kV for a 4 MeV beam.

The beam can be concentrated to a diameter of a few millimetres immediate above the analyser, by adjusting the extractor voltage applied to the ion source and the voltage applied to a focus electrode at the top of the tube. By observing the beam width at two points, some 7 ft apart, it has been found that the beam, when finally concentrated, is approximately parallel, which is the required condition for entrance to the analyser. Under these conditions up to 50% of the incident beam can be passed through the analyser with an energy resolution of 1 part in 1 200.

A deflecting magnet is normally used after the analyser. This bends the beam through $26 \cdot 8°$, yielding a horizontal beam; it also separates the protons from unwanted mass components of the beam. The design of the magnet poles and their distance from the exit slit of the analyser are such that the divergent beam entering the magnet is rendered parallel in the vertical plane. It diverges at an angle of 0·01 rad in the horizontal plane, but by using a very simple form of strong focusing it can be concentrated into an almost square spot of area less than 1 cm^2, at a target 12 ft beyond the magnet.

Proton currents up to 20 μA have been obtained after deflection by the analyser and magnet. The loss of particles of the required mass in passing through the magnet was estimated not to exceed 20%.

Mass analysis has shown a fairly poor proton content in the beam with H^+, H_2^+ and H_3^+ ions in the ratio 60 : 10 : 30.

The poor proton content, compared with a figure of 85% obtained from the source on test, may in part be due to the formation of neutral particles in and below the source canal as reported by Firth and Chick,[16] but it is also strongly dependent on the cleanliness of the source envelope and components.

The targets are in the form of thin film evaporated on to a suitable backing. Carbon deposition during irradiation is minimized by heating the target to 200° C and placing a liquid-air cold trap near it.

To establish fixed points in the high-voltage scale, the strong resonance reported by Herb at a proton energy of 873·5 keV in the reaction $F^{19}(p, \gamma)Ne^{20}$ was investigated and was redetermined as 874·7 keV with a standard deviation of 0·5 keV, while its width at half maximum yield was 5 ± 1 keV. The linearity of the analyser was checked by observing the same resonance with the mass-2 and mass-3 components and the discrepancy between the results was found to be 5 parts in 10^4.

The 2·57 MeV resonance in the reaction $Be^9(p, \gamma)B^{10}$ was determined as 2·567 MeV with a standard deviation of 0·002 MeV while its width at half maximum yield was 41 ± 3 keV. The yield curve is shown in Fig. 15.

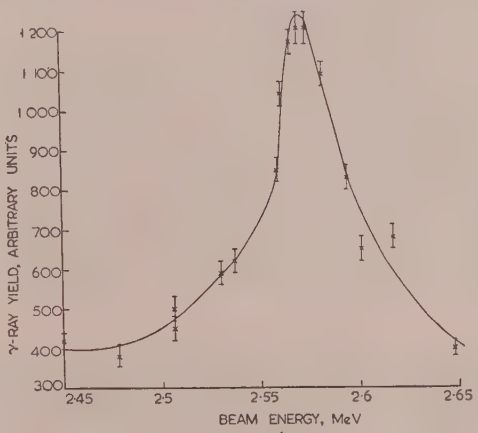


Fig. 15.—Yield curves for $Be^9(p, \gamma)B^{10}$ resonance at 2·567 MeV.

It was of interest to check the earlier method of generator-voltage estimation against results obtained using the electrostatic analyser. For analyser beam-energy measurements of 1·50 MeV and 2·85 MeV, values of top terminal potential of 1·49 MeV and 2·87 MeV were obtained by measuring currents in the machine, and using the equivalent circuit of Fig. 13.

(10) CONCLUSION

The accelerator has already been brought into use successfully as a nuclear research tool, producing beams of protons or deuterons up to 3 MeV energy with high energy homogeneity over long periods. For this purpose positive-ion suppression at the top of the tube was sacrificed to obtain reasonable beam focusing. A major redevelopment of the r.f. ion source and focusing system is being undertaken to obtain a well-focussed ion beam throughout the energy range in the presence of a positive-ion-suppressing field, with a single focusing control which is independent of ion source parameters. At present the

focusing is very sensitive to the extractor potential and r.f. oscillator power, and above 2·5 MV it has been impossible to obtain a well-concentrated beam.

Although the tube has usually been the limiting factor in voltage performance, another frequent source of trouble is surface flashover down the belt. Recently, negative charging of the downgoing run of the belt has been tried, the necessary voltage being obtained from a 100-megohm resistance in the lead from the stripping comb. This appeared to reduce the tendency for belt flashover, but further confirmation of this is needed, and also of its effect on the voltage stability of the generator.

One would expect that an initial charge density must be placed on the belt before the upper stripping comb can begin to operate, and that only the charge in excess of this can be stripped off. There is evidence that the initial charge density may be several times the useful density stripped off. The belt therefore carries a useless circulating charge causing unnecessarily high electric fields and spraying difficulties. Attempts have been made to measure the circulating charge, but it would appear worth while to use a separate voltage source in the terminal to improve stripping.

(11) ACKNOWLEDGMENTS

The authors wish to record the considerable assistance given by other electrostatic generator groups, in particular by Mr. J. W. Boag and the M.R.C. group, and Mr. R. L. Fortescue and the A.E.R.E. group, both in making available their own experience and in many helpful discussions.

Thanks are also due to the authors' many colleagues at Aldermaston, in particular to Dr. S. E. Hunt, for his work on the electrostatic analyser, and Messrs. B. Millar and J. L. W. Churchill, for their assistance in much detailed circuit design.

The authors also wish to record the valuable assistance given by the Research Department, Metropolitan-Vickers Electrical Co. Ltd., Manchester, in particular by Mr. J. Vipond, who was responsible for detailed mechanical design.

Thanks are due to Dr. T. E. Allibone, F.R.S., for permission to publish this paper, and for his detailed interest and active participation, to which the success of the project is very much due.

(12) REFERENCES

(1) Fortescue, R. L., and Hall, P. D.: "The High-Voltage Electrostatic Generator at the Atomic Energy Research Establishment," *Proceedings I.E.E.*, Paper No. 829, March, 1949 (**96**, Part I, p. 77).

(2) Boag, J. W.: "The Design of the Electrostatic Field in a Van de Graaff Generator," *ibid.*, Monograph No. 63, May, 1953 (**100**, Part IV, p. 63).

(3) Churchill, J. L. W., and Millar, B.: "Improvements in the Use of the Generating Voltmeter as a Servo Element," A.E.I. Research Report No. A.229, 1952.

(4) Richards, H. T., Smith, R. V., and Browne, C. P. "Proton-Neutron Reactions and Thresholds," *Physical Review*, 1950, **80**, p. 524.

(5) Chick, D. R., and Miranda, F. J.: "The Initial Performance of a Proton Accelerator Tube, Deterioration and its Causes," *Journal of Scientific Instruments*, 1952, **29**, p. 340.

(6) Millar, B.: "The Calculation of Voltage Surges in a Van de Graaff Generator," *British Journal of Applied Physics* (accepted for publication).

(7) Herzberg, G.: "The Spectra of Hydrogen," *Annalen der Physik*, 1927, **84**, p. 565.

(8) Thonemann, P. C., *et al.*: "The Performance of a New Radio-Frequency Source," *Proceedings of the Physical Society*, 1948, **61**, p. 483.

(9) Bayley, A. J., and Ward, A. G.: "A Positive Ion Source," *Canadian Journal of Research, A*, 1948, **26**, p. 69.

(10) Millar, B.: "A Remote Control and Telemetering System for a Van de Graaff Generator," *Electronic Engineering*, 1953, November and December, pp. 446 and 506.

(11) Harris, D. J.: Queen Mary College, London, Ph.D. Thesis, 1953.

(12) Trott, A. J. *et al.* (to be published).

(13) Bailey, R. (to be published).

(14) Chick, D. R., and Inall, E. K.: "Note on the Voltage Stabilization of an Electrostatic Generator," *British Journal of Applied Physics*, Supplement No. 2, 1953, p. 61.

(15) Hunt, S. E. (to be published).

(16) Firth, K., and Chick, D. R.: "The 'Screening' of Neutral Particles in a High Voltage Ion Accelerator Tube," *Journal of Scientific Instruments*, 1953, **30**, p. 117.

(17) Herlitz, J., see Dreyfus, L.: "Application of Theory of Conformal Representation to High Voltage Transformer Design," *Archiv für Elektrotechnik*, 1924, **13**, p. 123.

(13) APPENDICES

(13.1) Guard-Ring Profiles

Boag[2] has considered a conformal transformation giving a constant potential on a row of quasi-ellipses of the type required, taking account of the radial field but not the axial field. The transformation is

$$W = mz + \log_e\left[\frac{\sin(z+a)}{\sin(z-a)}\right] \quad . \quad . \quad . \quad (14)$$

where, as usual, $W = U + iV$ and $z = x + iy$.

The equation of the required ring profile is the $V = 0$ equipotential

$$\cos 2x = \cos 2a \cosh 2y - \frac{\sin 2a \sinh 2y}{\tan my} \quad . \quad . \quad (15)$$

The constants m and a determine the shape and spacing of the curves; x is directed down the stack and y radially outwards from an origin at the centre of one of the curves.

The corresponding field distortion factor to be applied to the radial field is found to be

$$f_h = 1 + \frac{2 \sin my_0}{m \sinh 2y_0} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where y_0 is the minor semi-axis obtained by solving equation (15) with $x = 0$.

The axial field distortion factor arising from the rings being at different potentials is estimated by calculating the radius of curvature ρ at the nose of the profile [$y = 0$ in eqn. (15)] and treating the field between rings as that due to infinite parallel cylinders of radius ρ and spacing equal to the gap between rings. This field distribution is well known and yields a field distortion factor

$$f_{ax} = \frac{\sinh p}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where

$$\sinh \tfrac{1}{2}p = \tfrac{1}{2}(g/\rho)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

g being the gap between the hoops. Expressions for both g and ρ can easily be deduced in terms of the parameters m and a from eqn. (15).

The choice of parameters $m = 5$ and $2a = 0{\cdot}65\pi$ yields the values $f_h = 1{\cdot}175$ and $f_{ax} = 1{\cdot}31$ quoted in the text.

(13.2) The Effect of Eccentricity

From a study of the well-known field pattern between parallel eccentric cylinders, the following expression was derived for the

distortion factor f_e on the inner cylinder due to a small eccentricity:

$$f_e = \sqrt{\left(\frac{S_0 + r_0}{S_0 - r_0}\right)} \frac{\log_\varepsilon (r_1/r_0)}{\text{arc cosh}\left(\frac{S_0}{r_0}\right) - \text{arc cosh}\left(\frac{S_1}{r_1}\right)} \quad . \quad (19)$$

where

$$S_0 = \frac{r_1^2 - r_0^2}{2b} - \frac{b}{2} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$S_1 = \frac{r_1^2 - r_0^2}{2b} + \frac{b}{2} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

r_0 and r_1 being the radii of the inner and outer cylinders respectively, and b the distance between their centres. This approximates to

$$f_e = 1 + \frac{2br_0}{r_1^2 - r_0^2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

when the eccentricity is small.

(13.3) The Intershield Design

The intershield skirt profile and its stress multiplying factor were studied as a two-dimensional problem using the Schwartz–Christoffel transformation, as modified by Herlitz[17] for a curved boundary. Taking $\zeta = \xi + j\eta$ as the potential function,

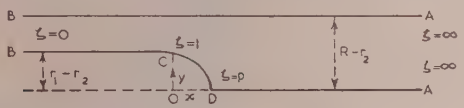


Fig. 16.—Diagram showing values of ζ used in the intershield theory.
BCD Intershield.
DA Lower stack column.
AB Tank wall.

and giving it the values shown in Fig. 16 at the points A, B, C and D, the transformation giving $\eta = 0$ over the boundary is

$$\frac{dz}{d\zeta} = C\left[\frac{\sqrt{(\zeta - 1)} + \lambda\sqrt{(\zeta - p)}}{\zeta\sqrt{(\zeta - p)}}\right] \quad . \quad . \quad (23)$$

or, in the integrated form,

$$z = C\left[\lambda \ln \zeta + \ln\left[\frac{\sqrt{(\zeta - 1)} + \sqrt{(\zeta - p)}}{\sqrt{(\zeta - 1)} - \sqrt{(\zeta - p)}}\right] - \frac{1}{\sqrt{p}} \ln\left\{\frac{\sqrt{[p(\zeta - 1)]} + \sqrt{(\zeta - p)}}{\sqrt{[p(\zeta - 1)]} - \sqrt{(\zeta - p)}}\right\}\right] \quad . \quad (24)$$

The constants C, λ, p are connected with the geometry of the figure by the three relations

$$C\left(1 - \frac{1}{\sqrt{p}}\right) = \frac{r_1 - r_2}{\pi} \quad . \quad . \quad . \quad . \quad (25)$$

$$C\lambda \ln p = k(r_1 - r_2) \quad . \quad . \quad . \quad . \quad (26)$$

where $k(r_1 - r_2)$ is the axial length OD of the skirt, and

$$C(\lambda + 1) = \frac{R - r_2}{\pi} \quad . \quad . \quad . \quad . \quad (27)$$

With the conditions $\eta = 0$ and $p > \xi > 1$, this can be split into real and imaginary parts giving the equations of the skirt profile CD:

$$x = C\lambda \ln \xi$$

$$y = 2C\left\{\text{arc tan}\sqrt{\left(\frac{p - \xi}{\xi - 1}\right)} - \frac{1}{\sqrt{p}} \text{arc tan}\sqrt{\left[\left(\frac{p - \xi}{p(\xi - 1)}\right)\right]}\right\} \quad . \quad . \quad . \quad . \quad (28)$$

To obtain the field strength, the further transformation $\zeta = \varepsilon^w$ is applied to equation (23) to put the conductors at different potentials, and the field is given, apart from a constant factor, by

$$\left|\frac{dW}{dz}\right| \propto \left|\frac{\sqrt{(\zeta - p)}}{\sqrt{(\zeta - 1)} + \lambda\sqrt{(\zeta - p)}}\right| = E \text{ (say)} \quad . \quad (29)$$

Over the region BC, $\eta = 0$ and $1 > \xi > 0$, the field is

$$E = \left[\frac{\sqrt{(p - \xi)}}{\sqrt{(1 - \xi)} + \lambda\sqrt{(p - \xi)}}\right] \quad . \quad . \quad . \quad (30)$$

At B where the field is uniform ($\xi = 0$), $E = \dfrac{\sqrt{p}}{1 + \lambda\sqrt{p}}$, while at C ($\xi = 1$), $E = \dfrac{1}{\lambda}$. The ratio of these gives the stress multiplying factor

$$f_s = 1 + \frac{1}{\lambda\sqrt{p}} = 1 + \frac{\ln p}{\pi k(\sqrt{p} - 1)} \quad . \quad . \quad . \quad (31)$$

The choice of values $k = 2{\cdot}9$ and $p = 41{\cdot}1$ leads to the factor $f_s = 1{\cdot}073$ quoted in the text.

[The discussion on the above paper will be found on page 152.]

537.534 : 621.317.7.082.72

Paper No. 1840 M
May 1955

AN ELECTROSTATIC ANALYSER FOR THE ABSOLUTE MEASUREMENT OF PROTON ENERGIES AND THE ESTABLISHMENT OF FIXED POINTS ON THE HIGH-VOLTAGE SCALE

By S. E. HUNT, Ph.D., D. P. R. PETRIE, M.Sc., K. FIRTH, B.Sc., and A. J. TROTT.

(*The paper was first received 8th October*, 1954, *and in revised form 28th January*, 1955. *It was published in May*, 1955, *and was read before the* MEASUREMENT AND CONTROL SECTION *25th October*, 1955.)

SUMMARY

An electrostatic analyser for measuring the energy of proton beams from an electrostatic generator is described. The energy measurements are absolute and are referred finally to the voltage of thermostatically controlled standard cells. The angle of deflection, 63·2°, was chosen to focus a parallel incident beam on to a fine exit slit. Target currents of over 20 μA were thus obtained with an energy resolution of 1 200.

The energy of a proton beam producing the lowest strong resonance in the $F^{19}(p\alpha\gamma)O^{16}$ reaction was determined as 874·5 ± 0·9 keV (absolute); the width of the resonance at half maximum yield was 5·0 ± 1·0 keV. Resonances in the $Be^{9}(p\gamma)B^{10}$ reaction were found for proton energies of 2·567 ± 0·003 MeV, 1·0837 ± 0·000 7 MeV and 989·5 ± 1·4 keV. The widths at half maximum yield of these resonances were 41 ± 3 keV, 3·8 ± 0·3 keV, and 91 ± 5 keV respectively.

(1) INTRODUCTION

The absolute measurement of proton energies to an accuracy of one part in a thousand in the energy range above 500 keV has been limited to the work of Herb *et al.*[1]

These workers used an absolute electrostatic analyser of deflecting angle 90° and measured the energies of proton beams for which $(p\gamma)$ resonances occurred. In particular their determinations of the first strong $(p\gamma)$ resonance in fluorine at 873·5 ± 0·9 keV has been used as a calibration point in the high-voltage scale by many later workers.[2,3] In order to provide an independent check on the high-voltage scale so established, an absolute electrostatic analyser of comparable accuracy has been used at this laboratory to measure and control the energy of proton beams from an electrostatic generator.

The analyser has a mean radius of 36 in and employs a deflecting angle of 63·2° in order to focus a parallel incident beam on to a fine exit slit. Currents of over 20 μA have been recorded on a target placed immediately after the exit slit, using an energy resolution of 1 200.

(2) SUMMARY OF THE THEORY

When a charged particle of mass m and charge e, having fallen through a potential V_1 is passed through a radial electric field, it follows from the simple Newtonian treatment that

$$V_1 = V_2 \frac{a_0}{2d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where V_2 is the potential difference between the electrodes producing the radial field, d is their separation, and a_0 is the mean radius of curvature of the particles. Relativistic treatment introduces a correction to the mean radius

$$a_r = a_0\left(1 - \frac{1}{2}\frac{eV}{m_0c^2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

It will be noted from the relativistic treatment that the mean radius a_r is no longer independent of the ratio e/m. A slight separation of the different mass components of the beam therefore takes place.

Hertzog[4] has shown that for a beam of particles describing a circular path of radius a in a radial electric field,

$$(l' - g)(l'' - g) = f^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where l' = Object distance measured from the entrance plane of the analyser.
l'' = Image distance measured from the exit plane of the analyser.
$g = a/\sqrt{2}\cot\theta\sqrt{2}$ = Distance of focal points from the entrance and exit planes of the analyser.
$f = a/\sqrt{2}\operatorname{cosec}\theta\sqrt{2}$ = Focal length.
θ = Angle of deflection.

The distances of the effective exit and entrance planes of the analyser from the ends of the electrodes were calculated from the formulae given by Herzog,[5] the slits having been fixed at a sufficient distance from the electrodes to prevent sparking.

The distance l'' between the effective exit plane and the plane of the exit slits was determined, and the angle of deflection was chosen so that g was equal to this. A parallel incident beam was then focused in the plane of the exit slits; the energy resolution of the system $W/\Delta W$ is given by $R = a/2w_2$, where w_2 is the width of the exit slit, W the beam energy and ΔW the spread in the beam energy. From Herzog's first-order theory, the resolution is independent of the width of the entrance slit. Second-order terms of Herzog's equations corresponding to aberrations of the system were considered and these gave the result that $R = 12a^2/5w_1^2$, where w_1 is the width of the entrance slit.

In practice the resolution of the instrument (1 200) was determined by the exit-slit width, and for this resolution the permissible entrance-slit width was much greater than the electrode separation.

(3) DESCRIPTION OF THE ANALYSER

The analyser is shown in Fig. 1. The geometric mean of the radii of the inner surfaces of the electrodes was 36·000 ± 0·000 5 in and their mean separation was 0·375 18 ± 0·000 04 in. The electrodes consist of low-carbon-steel bars of cross-section 2·5 in × 1·5 in. Each of these was fixed to the base-plate by four $\frac{3}{8}$ in diameter steel bolts passing through corrugated Perspex insulators, 5 in long, fitted with rubber gaskets as shown in Fig. 1. The bolts held the Perspex rigidly in compression against the metal surface. The base-plate was a mild-steel casting of 1·5 in thick with 4 in-thick webbing in order to prevent distortion. Both base-plate and bars were thoroughly stress-relieved and aged before machining. The faces of the gap were machined by mounting the cutting tools from a rigid cast-steel arm pivoted accurately at the centre of curvature of the electrodes. Hand

The authors are at the Research Laboratory of Associated Electrical Industries Ltd.

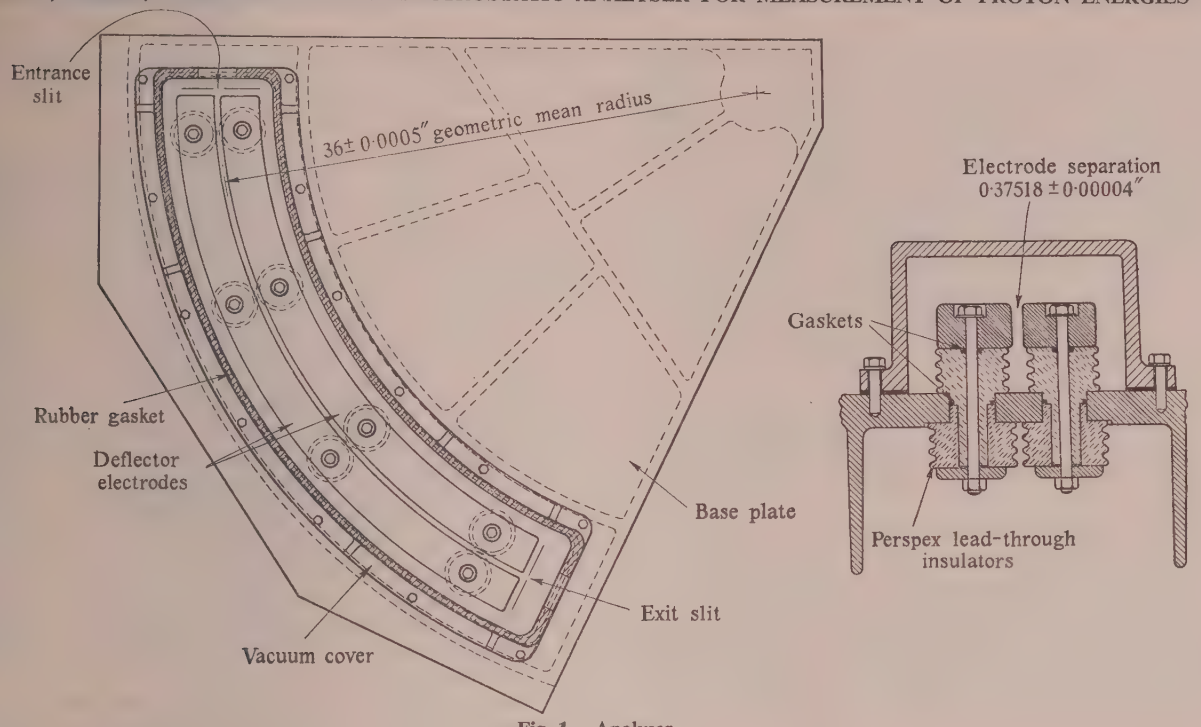


Fig. 1.—Analyser.

scraping tools were used until the constancy of the gap was 0·001 in, and a final cut was then taken from both faces simultaneously using an accurately lapped spiral end milling cutter. A constancy of rather better than 0·0002 in was thus obtained, and the mean separation was measured to an accuracy of 0·000 04 in.

The slit jaws consisted of flat copper sheets tipped with molybdenum over the area irradiated by the beam. They were accurately located by jigs fitting between the electrodes, and mounted on brackets fixed to the base-plate. They were insulated and the current falling on them was recorded.

The width of the entrance slit was slightly less than the electrode separation in order to reduce the bombardment of the electrodes by the beam, and the exit slit was set to a separation of 0·015 in.

The electrodes were enclosed in a vacuum-tight gunmetal cover sealed to the base-plate by a flat rubber gasket and fitted with four inspection windows. The analyser was mounted on a large turntable, and a flexible vacuum joint to the vacuum system of the Van de Graaff generator allowed it to be rotated through 180° about a vertical axis.

(3.1) Deflecting Magnet

The separation of the various mass components of the beam due to relativity effects depends on the beam energy but is of the order of the width of the exit slit; components of mixed mass are therefore allowed to pass through the analyser. In order to separate these a deflecting magnet was used. This had a deflecting angle of 26·8° and its focal length was equal to the distance between the exit slit of the analyser and the entrance plane of the magnet. A parallel horizontal exit beam was therefore produced.

(4) DEFLECTOR VOLTAGE SUPPLIES

The deflector voltage stabilizing circuit is shown schematically in Fig. 2, and will be reported fully elsewhere (Millar, Bailey and

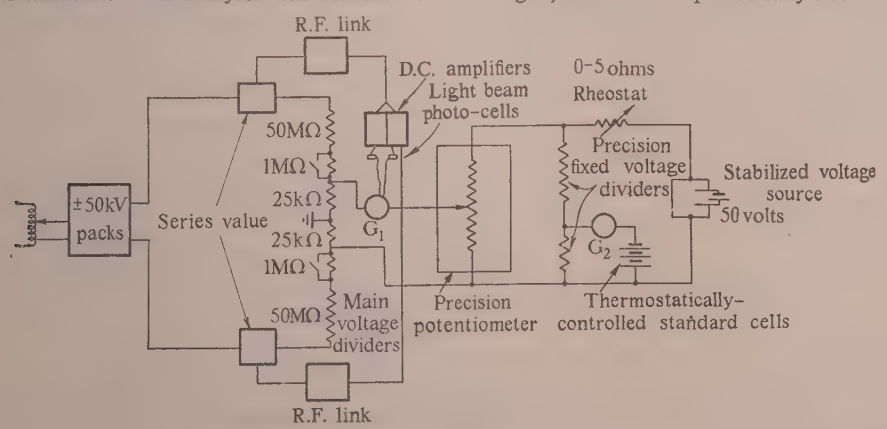


Fig. 2.—Stabilizing circuit for ±50 kV deflector supply.

Churchill[6]). The ultimate voltage reference was eight thermostatically controlled standard cells, connected via a galvanometer to a precision voltage divider. Any deflection of the galvanometer produced a signal via a pair of photocells which adjusted the 0·5-watt rheostat to maintain the voltage across the variable precision potentiometer constant. Similarly, any signal in a galvanometer connected between the tapping point of the potentiometer and the low-ratio arm of one of the high-voltage voltage dividers was made to alter the impedance of a valve in series with the output of the deflector power-pack to maintain this constant. The setting of the precision potentiometer therefore controlled the voltage applied to the electrodes, which was also symmetrical with respect to earth potential.

The main voltage dividers consisted of ten units of fifty, 100-kilohm wire-wound resistors mounted on Perspex frames, and the low-ratio arms were similar resistors of value 22 kilohms. Each frame was surrounded by a spinning maintained at the mean potential of the enclosed resistors by an auxiliary chain of carbon resistors, in order to minimize errors due to corona currents.

The voltage-divider ratios were measured by a method similar to that used by Rymer and Wright.[7]

A 1-megohm manganin wire-wound resistor of accuracy within $\pm$ 1 part in 10^4, which was normally short-circuited, was connected between the high- and low-ratio arms. In order to measure the ratios of the voltage dividers whilst they were subjected to the working voltage, the two dividers were connected in parallel to one polarity of the high-voltage source as shown in Fig. 3.

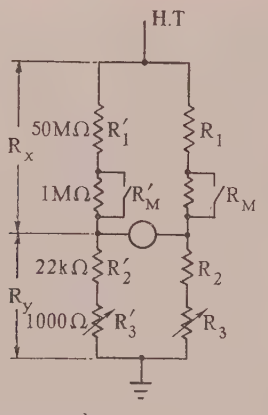


Fig. 3.—Circuit for voltage-divider ratio measurement.

The other end of each voltage divider was connected to earth through a precision 1 000-ohm variable resistor, and a galvanometer was connected across the two junction points between the 1-megohm and 22-kilohm units.

In the Wheatstone bridge so formed R_m was short-circuited, R_3 made equal to zero, and R_3' adjusted for balance.

Then
$$\frac{R_x}{R_y} = \frac{R_1}{R_2}$$

R_m was then put in the circuit and R_3 adjusted to retain the balance, whereupon

$$\frac{R_x}{R_y} = \frac{R_1 + R_m}{R_2 + R_3}$$

hence the voltage divider ratio is

$$\frac{R_1}{R_2} = \frac{R_m}{R_3}$$

Since both R_m and R_3 were accurately known, the voltage-divider ratio was found to a comparable accuracy. The process was repeated for each voltage divider using applied voltages of both polarities up to the maximum used on the electrodes.

(5) PERFORMANCE OF THE ANALYSER

Beams of energy up to 3·25 MeV have been passed through the analyser, and the maximum deflector voltages applied to the electrodes ($\pm$40 kV) were sufficient to deflect a 4 MeV beam. The electrodes showed a definite "conditioning" effect, initial sparking taking place for voltages as low as $\pm$10 kV even after careful cleaning. Occasional flashover was observed across the insulators inside the vacuum, and at one stage a slightly burnt track was found (this was removed by scraping), and the impedance in series with the electrodes was increased to 100 kilohms, which limited the current flowing at flashover to a safe value.

Meters were installed between the electrode and the high-voltage source, and currents of a few hundred microamperes were observed to flow on occasions. This is thought to be due to a charge-multiplication process occurring between the electrodes. The effect was worse when wide incident beams were used, owing to secondary emission from the entrance-slit jaws striking the top of the electrodes and initiating the process.

It was found possible to produce a parallel incident beam of cross-section rather less than the entrance-slit width, so that the current intercepted by the entrance slits was only a few microamperes. Under these conditions the current to the electrodes was about 20–30 μA.

Target currents of over 20 μA have been obtained, although for most of the $(p\gamma)$ resonance work reported here, beams of between 2 and 10 μA were used, to avoid damage to the targets and in some cases to avoid excessively high γ counting rates. Using no secondary-electron suppression, similar currents were recorded on the exit-slit jaws, and the beam width at the exit plane seen through the windows in the cover was about twice the exit-slit width. This could be explained by the splitting of the beam into its various mass components owing to relativity effects, but the focus would also be broadened by the use of a non-parallel or astigmatic incident beam or by any energy spread in the beam produced in the ion source.

Currents intercepted on the exit-slit jaws were used to control the voltage of the generator by varying the intensity of an electron beam up the accelerator tube.[6] The response time of the electronic circuits controlling this current was extremely fast; consequently the position of the beam on the exit slits was constant and very little fluctuation in target currents was noted.

(6) RESONANCE MEASUREMENTS

In order to establish fixed points on the high-voltage scale it was decided in initial irradiations to determine absolutely the proton energy at which a few large $(p\gamma)$ resonances occurred. An obvious choice for this purpose was the first strong resonance occurring in the $F^{19}(p\alpha\gamma)O^{16}$ reaction, previously measured by Herb as having a peak γ-ray yield for a proton energy of 873·5 keV, and as additional fixed points resonances in the $Be^9(p\gamma)B^{10}$ reaction were chosen.

The experimental procedure was to prepare metallic beryllium or calcium fluoride targets by evaporation *in vacuo* on to carefully cleaned and polished copper discs. These were irradiated by placing them either directly after the analyser or, when the mixed-mass components of the beam could cause confusion, after the deflecting magnet. The γ-rays emitted were detected by means of either a Geiger–Müller or a scintillation counter placed immediately behind the target. The γ-ray yield was plotted as a function of beam energy for a constant beam charge on the target. It had previously been found necessary to maintain the target material at a temperature of about 200° C and to have

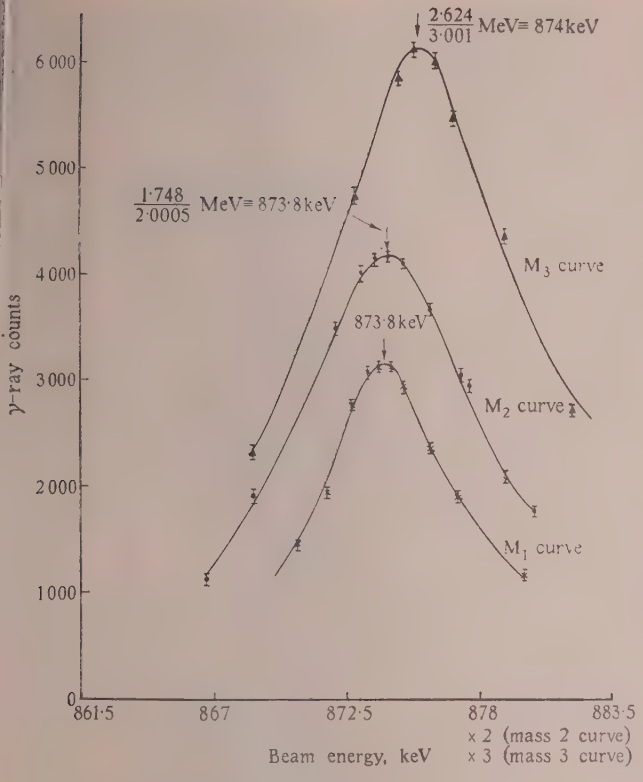


Fig. 4.—The 874·5 keV $F^{19}(p\alpha\gamma)O^{16}$ resonance.
Observed using Mass 1, Mass 2 and Mass 3 components of the beam.

a liquid-air trap in its immediate vicinity in order to prevent the deposition of carbon on the target surface.[8]

The thickness in energy units of the targets used was estimated from the geometry of the evaporation process and the stopping power of the materials.

In order to check the linearity of the voltage scale the fluorine resonance was observed using the mass 1, mass 2 and mass 3 components of the beam. Singly-charged diatomic hydrogen molecules split up into two protons and an electron on hitting the target material, each proton having $[1/(2 + 1/1\,840)]$ of the initial energy, and similarly each proton from a singly-charged triatomic molecule has an energy equal to $[1/(3 + 2/1\,840)]$ of the initial energy, thus if Ve is the energy corresponding to the resonance observed using the proton beam, resonances will be observed at $(2 + 1/1\,840)Ve$ and $(3 + 2/1\,840)Ve$ using mass 2 and mass 3 beams. The results obtained using these three mass components of the beam are shown in Fig. 4. The discrepancy between the three results was $0·5/10^3$, which is of the order of the variations observed even when using beams of the same mass component.

The mean value obtained for the fluorine resonance was 874·5 keV (absolute), the standard deviation of twelve separate determinations being 0·5 keV. Other resonances in fluorine have been measured absolutely and will be reported at a later date.

The mean value of five determinations of the large beryllium resonance was 2·567 MeV, the standard deviation being 0·002 MeV. Its width at one-half maximum yield was 41 ±3·0 keV. Other resonances were observed in beryllium at 989·5 keV (standard deviation, 1·0 keV) and 1083·7 (standard deviation, 0·3 keV) of half-widths 91 ± 3 keV and 3·8 ± 0·5 keV respectively. The beryllium resonances are shown in Fig. 5. In all resonance values quoted, correction has been made for the relativity effect, which was −0·3 keV for the 874·7 keV resonance and −3·5 keV for the 2·567 MeV resonance.

(7) ESTIMATION OF OVERALL ACCURACY

There are eight sources of error which could affect the overall accuracy of the results. These will be briefly dealt with in turn.

(7.1) Measurement of the Ratio $a/2d$

The separation of the electrodes was measured by a dynamic air-flow method[8] capable of detecting variations of 0·000 01 in. The accuracy of the measurement therefore became that of the standard separation gauge against which the gauge head was calibrated. This was 0·000 04 in, and measurements taken during an interval of several months varied by less than ±0·000 03 in.

The radius measurement was made at the time of manufacture and was 36·000 in ± 0·000 5 in. Although measurements of gap and mean radius were made at 20° C, at which temperature standard gauges were accurate, the ratio $a/2d$ is essentially independent of ambient-temperature variation, since the temperature coefficients of the base-plate and electrode material were very nearly equal. Local heating of the electrodes could reduce the gap, but because of the very massive construction of the electrodes this is thought to be unlikely.

Errors in the estimation of $a/2d$ are therefore estimated to be less than 1 part in 10^4.

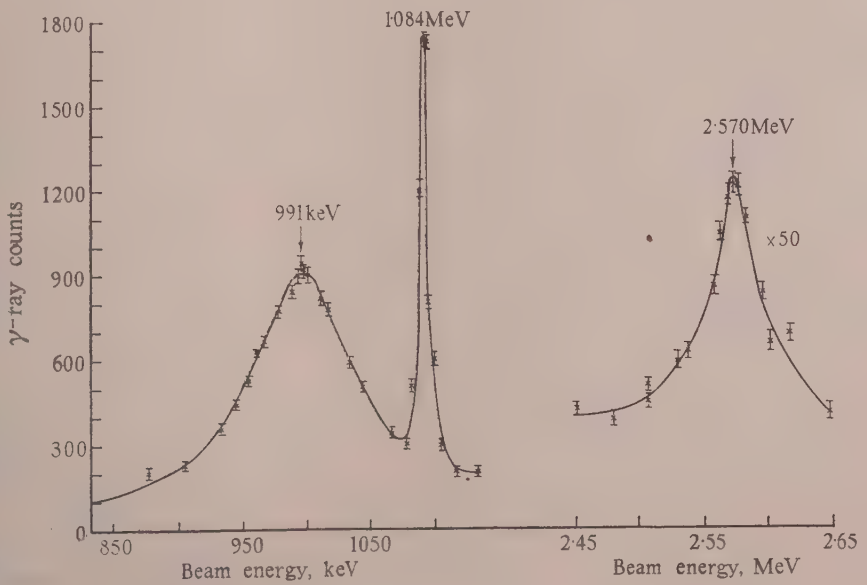


Fig. 5.—Resonances observed in the $Be^9(p\gamma)B^{10}$ reaction.

(7.2) Positioning of the Slits

Herzog's calculations for the determination of effective entrance and exit planes have been applied.[5] For the exit slits it has been assumed that the beam travels in a circular path until it reaches the exit plane and then continues tangentially to the exit slit. In order to allow for this the exit slit was off-set from the mean radius by a small amount y given by

$$y = \frac{(l'' - S)^2}{2a}$$

where S is the distance of the exit plane from the ends of the electrodes. This was 0·001 3 in, so that the correction was very small compared with the mean radius. The radial error in setting the exit-slit jaws from the inner faces of the electrodes using a precision jig was less than ±0·000 5 in, so that the mean radius of the beam is thought to be equal to the geometric mean radius of the electrodes within these limits, i.e. within ±0·5 part in 10^5.

The positioning of the entrance slit is important in that it determines the effective entrance plane of the analyser and consequently the position of the focus. It is also desirable that this entrance plane should be horizontal.

(7.3) Orthogonality of the Entrance Plane and Incident Beam

From the optical analogy, the internal displacement of the beam at the exit plane produced by an inclination α (rad) of the incident beam to the vertical is given by $b'' = \alpha f$. Since $f = a/\sqrt{2}\ \mathrm{cosec}\ \theta\sqrt{2}$, this produces a fractional error in radius measurement $b''/a = \alpha'/\sqrt{2}$. The accelerator tube was built accurately vertical; it was therefore assumed that the incident beam to the analyser was approximately vertical. A spirit level was mounted on the analyser base-plate in a plane parallel to the calculated entrance plane of the analyser, and the analyser was accurately levelled before resonance measurements were made; it was then rotated through an angle of 180° about a vertical axis using a large turntable, again accurately levelled, and the measurements were repeated. Assuming an initial inclination α between the incident beam and the normal to the entrance plane, in the second set of measurements this became $-\alpha$. The difference between sets of measurements taken in the two analyser positions was very close to one part in 10^3 throughout the voltage range, indicating a value for α of about 0·04°. The resonance values quoted are the mean values of measurements taken with the analyser in the two positions, so that the systematic error was cancelled out. There remains an error of ±2 parts in 10^4 due to the possible error in levelling the analyser. As this was done on numerous occasions, this error can be regarded as random throughout the results.

(7.4) Homogeneity of Electric Field

The depth of the electrodes was 1·25 in, and the beam was restricted to a central portion of the gap of depth 0·375 in. The field within this depth has been calculated to be homogeneous to within 5 parts in 10^5.

(7.5) Stray Magnetic Fields

The effect of stray magnetic fields perpendicular to the beam and to the base-plate of the analyser is to modify eqn. (1) as follows:

$$V_1 = V_2 \frac{a_0}{2d\left(1 - \frac{Ha}{2c}\Big/\frac{2e}{mV}\right)}$$

Gaussian units being used.

Radial magnetic fields merely produce slight displacements of the beam within the depth of the electrodes, and do not affect the beam-energy measurements.

The perpendicular magnetic field was investigated using a search coil and sensitive ballistic galvanometer, and was found to have a mean value of 0·15 gauss in the direction out from the base-plate. This introduced a correction of one part in 10^4 for an incident beam of 1 MeV.

The field was largely due to the magnetization of the deflector electrodes and baseplate and was reasonably reproducible. The error associated with stray magnetic fields is therefore estimated to be less than ±1 part in 10^4 even for the lower-resonance measurements.

The stray field produced by the deflector magnet was less than 0·2 gauss at the end of the analyser near the magnet when this was set for a beam of energy 1 MeV; this would produce a constant error of about 1 part in 10^4 over the whole energy range. To check this, resonance determinations were made placing the target directly after the analyser. The magnet was in position and no difference in the resonance peaks could be noticed when it was energized by the current appropriate to the beam energy used.

(7.6) Errors due to Uncertainty in Voltages applied to the Electrodes

From the schematic (Fig. 2) it will be seen that the accuracy of the voltage measurements depends essentially upon the constancy of the standard cells, the fixed voltage-divider ratio, the high voltage-divider ratio and the accuracy of the precision potentiometer.

The standard cells were of miniature type and were thermostatically controlled at a temperature of 35° C, and the current drain was essentially zero.

The fixed voltage-divider was a precision instrument of accuracy within ±1 part in 10^4.

The high voltage-divider ratios were measured by the method outlined in Section 5, over a range of applied voltage from 5 kV to 40 kV. The sensitivity of measurements was proportional to the applied voltage and was about ±1 part in 10^4 for an applied voltage of 10 kV. No systematic variation in ratio was noted with applied voltage, so the ratios were constant with voltage to 2 parts in 10^4 or better. The accuracy of all ratio determinations depended on the accuracy of the standard 1-megohm units R_m and the variable resistance boxes R_3. The former were made up of manganin coils accurate to 1 part in 10^4, and the latter were accurate to ±2 parts in 10^4. The probable error associated with the ratio measurements is therefore estimated to be ±3 parts in 10^4. Ratio measurements were made with the leads of the stabilizing loop connected to the voltage dividers; any shunting effect due to finite impedance of these was found to be negligible.

Initial resonance measurements were made using a precision potentiometer which was subsequently found to have errors considerably in excess of the quoted accuracy (1 part in 10^4) owing to a failure in one of the lower-decade coils. The potentiometer was calibrated against a similar instrument recently standardized by the National Physical Laboratory; nevertheless it was necessary to discard many earlier resonance measurements in the low-energy range, since it was not definitely known when the fault occurred. In the energy range above one million volts the correction was less and has been applied for measurements taken over a short period before the fault was located. The lower-energy resonance measurements were repeated using a different instrument, the errors of which were less than 0·5 part in 10^4 compared with the N.P.L. standardized potentiometer. This instrument was also used to check the fixed voltage-divider ratio, variable resistance boxes (R_3) and the 1-megohm units R_m.

It was also used to check the voltage of the standard cells against other standard cells of a larger type maintained at 20° C.

The overall uncertainty in the measurement of the deflector voltages is estimated to be rather less than 5 parts in 10^4.

(7.7) Errors associated with Target Thickness

The thicknesses of targets, ΔW, used for the fluorine resonance was 1 or 2 keV, and errors in estimating this thickness from the stopping-power curves are probably about 10%. The observed resonance peak occurs at a beam energy of

$$W_R = W_0 + \Delta W/2 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where W_0 is the true resonance energy. Uncertainties due to target thickness in the fluorine resonance measurements are estimated to be less than 0·1 keV, or 1 part in 10^4 of the final energy determination.

Targets of thickness 2, 5 and 7 keV were used for the beryllium resonances. Uncertainties of up to 0·35 keV exist for these measurements. Several irradiations were also performed with 30 keV-thick targets; the results, after correction for target thickness, were in good agreement with those obtained using the thinner targets, and they can be taken as a confirmation of the accuracy of the thickness estimates.

(7.8) Errors associated with Counting Statistics

The yield of all the resonances was high, as will be seen from Figs. 4 and 5, and the accuracy with which the peak could be estimated depended largely upon the width of the resonances. This varied from $\pm 0{\cdot}3$ keV for the narrow resonances to ± 3 keV for the broad, 989·5 keV beryllium resonance.

These are the uncertainties in reading from a single curve, and should be purely random. The mean values of several curves should consequently be more accurate.

(8) DISCUSSION OF RESULTS

Of the errors previously discussed, those in Sections 7.3, 7.7, 7.8, and to a large extent Section 7.6, are random for different irradiations made over a period of some months. The standard deviations of the resonance measurements themselves can therefore be used as a check on the random components of these errors. The standard deviation of the fluorine results, $\pm 0{\cdot}5$ keV, is rather less than the root of the sum of the squares of estimated random errors, which confirms that no large random source of error has been omitted. The same is true for the beryllium resonances.

The total systematic error is estimated to be about ± 4 parts in 10^4. By adding the estimated systematic error to the standard deviation of several resonance determinations, the results become $874{\cdot}5 \pm 0{\cdot}9$ keV for the $F^{19}(p\alpha\gamma)O^{16}$ resonance, and $2{\cdot}567 \pm 0{\cdot}003$ MeV, $1{\cdot}083 \pm 0{\cdot}000\,7$ MeV and $989{\cdot}5 \pm 1{\cdot}4$ keV for the three $Be^9(p\gamma)B^{10}$ resonances. A voltage scale based on the authors' determinations of the fluorine resonances would be 1 part in 10^3 higher than that based on Herb's result. The discrepancy is just within the combined limits of error of the two determinations.

No direct absolute determination of the $Be^9(p\gamma)B^{10}$ resonances has been made previously. Fowler, Lauritsen and Lauritsen[9] obtained values of 988 and 1 077 keV respectively for the lower resonances, using an electrostatic analyser calibrated from the $Li^7(p\gamma)Be^8$ resonance, taken as 440 keV. If more recent values are taken for the lithium resonance,[3,8] Lauritsen's values would become 991·5 keV and 1·081 MeV.

Hahn *et al.*[11] have obtained a value of $2{\cdot}565 \pm 0{\cdot}01$ MeV for the higher beryllium resonance using a magnetic deflector calibrated against Herb's determination of the fluorine resonance. Hahn's result is 8 parts in 10^4 less than that obtained in the present work, but the errors associated with his determinations are comparatively large.

Richards *et al.*[10] obtained a value of 2·56 MeV based on Herb's determination of the lithium (pn) threshold at $1{\cdot}882 \pm 0{\cdot}002$ MeV. From this our voltage scale would be 2·6 parts in 10^3 higher than that of Herb; errors associated with Richards's results are not quoted. Our result is 1·7 parts in 10^3 higher than the mean of that obtained by Hahn and Richards; this may be significant in confirming a disagreement of this order between our voltage scale and that established by Herb.

(9) ACKNOWLEDGMENTS

The authors are pleased to acknowledge the assistance of many colleagues in this work. In particular Messrs. D. R. Chick, B. Millar and J. L. W. Churchill were responsible for the voltage stabilization of the generator, and Messrs. R. Bailey and I. Collip for stabilizing the deflector voltage.

Prof. Rotblat has contributed many helpful suggestions in the course of this work, and the authors would also like to thank members of the workshop staff of the laboratory, together with Messrs. Wilson and Kyle, who shared responsibility for the manufacture of the analyser, and Mr. P. J. Young, who assisted in the experimental work.

Thanks are also due to Dr. T. E. Allibone, F.R.S., for permission to publish the paper.

(10) REFERENCES

(1) Herb, R. G., Snowden, S. C., and Sala, O.: "Absolute Voltage Determination of Three Nuclear Reactions," *Physical Review*, 1949, **75**, p. 264.

(2) Willard, H. B., Bair, J. K., Kington, J. D., Hahn, T. M., Snyder, C. W., and Green, F. P.: "The Yield of Gamma-rays and Neutrons from the Proton Bombardment of Fluorine," *ibid.*, 1952, **85**, p. 849.

(3) Fowler, W. A., and Lauritsen, C. C.: "Beryllium-proton Reactions and Scattering," *ibid.*, 1949, **76**, p. 314.

(4) Herzog, R.: "Ionic- and Electro-optical Cylindrical Lenses," *Zeitschrift für Physik*, 1934, **89**, p. 447.

(5) Herzog, R.: "Calculation of a Stray Field of a Condenser, the Field being limited by a Disc," *ibid.*, 1935, **97**, p. 596.

(6) Millar, B., Bailey, R., and Churchill, J. L. W.: "The Stabilization of a 4 MeV Electrostatic Accelerator." (In course of preparation.)

(7) Rymer, T. B., and Wright, K. H. R.: "Potentiometer Circuit for Measurement of High Potential," *Journal of Scientific Instruments*, 1952, **29**, p. 139.

(8) Hunt, S. E.: "Some Absolute Determinations of Proton-gamma Resonances in the Energy Range below 500 keV," *Proceedings of the Physical Society*, 1952 A, **65**, p. 982.

(9) Fowler, W. A., Lauritsen, C. C., and Lauritsen, T.: "Gamma Radiation from Excited States of Light Nuclei," *Reviews of Modern Physics*, 1948, **20**, p. 236.

(10) Richards, H. T., Smith, R. V., and Browne, C. P.: "Proton-Neutron Reactions and Thresholds," *Physical Review*, 1950, **80**, p. 524.

(11) Hahn, T. M., Snyder, C. W., Willard, H. B., Bair, J. K., Klema, E. D., Kington, J. D., and Green, F. P.: "Neutrons and Gamma-rays from the Proton Bombardment of Beryllium," *ibid.*, 1952, **85**, p. 934.

[The discussion on the above paper will be found overleaf.]

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL SECTION, 25TH OCTOBER, 1955

Dr. J. W. Boag: The authors have chosen electrode shapes and sizes to give minimum field strength for a given voltage. There is, of course, no completely satisfactory optimum design: the inevitable engineering compromise appears when the authors choose for the intershield potential difference the mean of the optimum voltages for cylindrical and spherical geometry. A very slight improvement is, I believe, possible if the spherical end of the intershield is not concentric with the spherical end of the top terminal but has a slightly greater axial spacing.

It is well known that in very inhomogeneous fields Paschen's law is not obeyed, and the sparking voltage, after passing through a maximum, may even decrease again with increasing pressure. The electrode geometry in a Van de Graaff generator is relatively smooth, however, and it is surprising that such an effect should appear in this and many other machines at only 10 atmospheres. Not very much work has been done on gas breakdown at voltages above 1 MeV, and it would be of considerable engineering interest to get some data on breakdown as a function of both gap length and gas pressure. This would be possible if a movable electrode at tank potential were fitted temporarily. It may be, of course, that breakdown is initiated inside the column—perhaps at the spark gaps between plates or along the belt itself.

I have always considered that the main purpose of the belt "guide bars" was to control the potential difference both across and along the belt, so as to hold the maximum charge density on the belt. To perform this function metal bars would seem to be necessary, and I should like to hear a little more about the authors' reasons for preferring Perspex and their interpretation of how it works. If the bars acquire a surface charge, this would almost certainly come from the belt and be of the same sign. I have not been able to see how a charge distribution of this kind would lead to a smooth gradation of potential along the belt, and one would not expect it to be a very stable arrangement in any case. On the other hand, with the relatively wide plate spacing used in this machine, metal bars would give a rather high local stress concentration. There might be a case for using metal bars of elliptical section.

The most critical component in all Van de Graaff machines is the accelerating tube, and I hope that at some future time the authors will give us a fuller account of their experience with the several designs of tube which they have now tested. I am not familiar with the theory of vacuum breakdown put forward by Harris, but it has to be remembered that, in an accelerating tube of the great length used in this machine, the chance of a neutral particle liberated at the bottom reaching the top again is only about 1 in 10^4.

Dr. H. R. Allan: A few points have arisen in our experience with a 2 MeV Van de Graaff generator at Imperial College, on which the authors may be able to comment.

Although we also use Freon to improve the insulation of the tank filling, we are worried about the corrosive action of free fluorine that may be liberated because of sparking. For this reason, and also for economy, we have followed the American practice of using a mixture of nitrogen and carbon dioxide, which is said to give an appreciable improvement over nitrogen alone. But we still add Freon when we want the highest voltages available from the generator.

Voltage stabilization by an electron beam necessarily adds to the X-ray intensity near the generator. When precise measurements with counters are being taken, such X-rays can be a serious trouble. Another drawback is that the X-rays produce ionization in the tank filling, which may make stabilization more difficult.

Oil diffusion pumps have been a cause of trouble in the Imperial College generator, and similar experience has been reported from America. After a prolonged period of use we found that the focusing electrode and its insulating supports were covered with a black incrustation of (presumably) cracked oil vapour. This prevented the focusing voltage from being properly established, and the beam failed to focus. The electrodes had to be cleaned, and we now use mercury pumps instead of oil ones. Our oil pumps had only a water-cooled baffle (not CO_2 or liquid air), but it seemed best to adopt mercury pumps, and so reduce contamination of the accelerating tube, and of targets, as far as possible.

The argument for having a wide accelerating tube when one wants large beams is, presumably, to have adequate pumping speed to remove the gas that emerges from the ion source canal with the beam. It should, however, be possible to obtain adequate beams with quite a small gas flow by careful design of the ion source itself.

Differential pumping is sometimes mentioned as a solution to this problem, but I understand that it is now held to be unsatisfactory. The differential pumping tube never gets properly outgassed (because it does not carry a beam) and so tends to be the limiting factor on the voltage attained by the generator.

Mr. D. Moore: The Medical Research Council's 2 MeV Van de Graaff generator at Hammersmith Hospital was designed for use in biological research, where the accent is on high currents and not on precision voltage stability. The demands are unusual since an output of X-rays, electrons and neutrons is required. These criteria affect the detail design and require, for example, a wide accelerating tube and an efficient belt-charging system.

We have a system which avoids circulating belt currents. With a set of three spray combs, in the high-voltage terminal, it is arranged that the down-going side of the belt is recharged with the opposite polarity to the up-going side. We have found it necessary to make one of the comb gaps adjustable from the control desk. An extra comb near the lower pulley is used to collect the charge carried by the down-belt run, a valuable feature of which is to provide a measure of the current at the control desk. This arrangement facilitates stable operation on both polarities with either high or low currents.

A new accelerating tube was designed in 1952 under Dr. Boag's guidance, with a symmetrical electrode system. The porcelain sections were ground flat and cemented to stainless-steel rings on which the electrodes were riveted. The pitch was $2\frac{9}{16}$ in and the tube length 7 ft 6 in. A mass-analysed ion source was incorporated at this time, designed to produce a 0·5 mA collimated beam of deuterons. The tube was conditioned to 2 MV positive and 1·9 MV negative, and we were able to operate at currents up to 350 μA of deuterons. Shortly afterwards we ran into an unexpected difficulty. The maximum voltage the tube would hold fell progressively to 500 kV, at which value there was a high unfocused electron current in the tube and a substantial X-ray emission. We believe that this was due to electron multiplication on the inner surface of the electrodes and that there must have been some change in surface conditions, possibly due to contamination by oil vapour. This particular design may lend itself to electron multiplication because of the considerable area of electrode surface almost tangental to a normal beam, which may serve to make this an efficient electron multiplier tube.

In an endeavour to overcome this trouble, diaphragms were inserted on every third electrode to waist the tube from 6 in at the ends to $4\frac{1}{2}$ in at the centre. Subsequently tube performance has improved, with the result that the machine is frequently run at 2·25 MV on positive polarity with beam currents of 350 μA

during neutron irradiations for biological experiments; the maximum voltage negative is at least 2 MV. We believe that the performance of the tube with diaphragms is less affected by traces of oil vapour.

A vacuum failure occurred in which the belt driving a mechanical pump broke during a weekend and the safety devices failed to operate. Hot diffusion-pump oil was distilled into the tube. After the pumps alone had been cleaned and repaired the tube was reconditioned, its previous performance being quickly restored. Our vacuum failures have been limited to this mishap; this is perhaps of interest, since we exercised less care in preparing the components of the accelerating tube than was, I believe, taken at Aldermaston. The stainless-steel rings were trepanned from stock descaled sheet of $\frac{1}{16}$ in thickness. Without further machining these were cemented with vinyl acetate to lapped porcelain rings. The joints so formed have been entirely leak-free.

Recently a new set of column voltage-divider resistors has been installed to reduce the current drain from the terminal; this has improved the overall efficiency of the machine, enabling a target current of 350 μA to be obtained with 350 μA of spray current.

Our metering system does not permit the comprehensive current measurements the authors made with their machine, but typical performance figures for the Van de Graaff generator at 2·2 MV on positive polarity with a deuteron beam of 350 μA would be: spray 350 μA; down-belt current 125 μA; voltage-divider resistance current 110 μA, with a vacuum of 2×10^{-5} mm Hg.

Mr. P. Howard-Flanders: In common with many others, the authors have found that the machine will produce a higher voltage than the tube will withstand. The mechanism of tube breakdown is not at present understood, and the development of high-voltage tubes has been almost entirely empirical. Experience with new designs, such as those reported, is particularly valuable in extending present knowledge, and may enable us to find out what factors are important in tube construction.

The paper makes a contribution to the controversy as to whether mercury diffusion pumps are essential for good tube performance. The authors have shown that a performance of 3·8 MeV can be achieved in a 9 ft tube when oil diffusion pumps with a refrigerated baffle are used. This performance is good enough to suggest that there cannot be a very great difference between a well-trapped oil system and one using mercury pumps.

As the problem of focusing the ion beam has been referred to, some of our experiences with the Medical Research Council's 2 MeV machine at Hammersmith Hospital may be of interest. We have had to cater for an unusually wide range of operating conditions, since the machine is used for accelerating either electrons or positive ions. It is arranged that the electric field is low in the first short section of accelerating tube between the ion source and lens, but is high and uniform along the remaining length. The change of electric field acts as a powerful lens and focuses particles from the ion source or filament on to the target. Voltages for all the electrodes are obtained from the resistance voltage divider, so that no additional power supplies are needed in the high-voltage terminal. The machine can be operated on either positive or negative polarity by means of controls placed outside the pressure vessel. Moreover, the system remains in focus for a wide range of operating voltages or injection energy without adjustment. In our machine we have found that while operating at a current of 350 μA of deuterons at 2 MeV, the magnification of the system is 2·5 at the end of a target extension tube 1·6 times the length of the accelerating tube.

Mr. R. L. Fortescue: I should like to ask to what extent the authors think this use of maximum field strength is justified. Two figures are given for field strength which has been satisfactorily withstood—160 kV/cm and 210 kV/cm.

It would be interesting to know what it is that fails above these field strengths—whether it is, in fact, a failure at a particular local maximum field strength and whether there is any evidence that one can use these figures for longer or shorter gaps.

In respect of Figs. 7 and 8 on current performance, it is not clear to me whether there is Freon present or nitrogen only.

It is stated that a different sort of comb was tried in order to improve the stripping, but there are no comments on what happened. It would be interesting to know whether it was successful or whether it is thought necessary to use a spray voltage in the top terminal to get satisfactory stripping.

The authors also say that the time lags are different between the gas breakdown and the tube breakdown, the tube breakdown being very much faster. If we could be told more about how this is measured the use of the information in this paper for other purposes would be easier.

Dr. Boag referred to some work at Queen Mary College and argued that for the breakdown of a tube of this length it is rather hard to see how neutral particles could come into the question. I think one point is that there are a lot of other metal objects in the tube besides its top end. One does not necessarily have to visualize a neutral particle going non-stop from the bottom to the top, particularly since the voltage may not disappear completely on these occasions.

Dr. A. Nemet: How do the authors calibrate their generating voltmeter? The indication of this instrument depends on the geometry of the electric field and cannot be calculated with satisfactory accuracy in most cases.

Secondly, how large is a 50-megohm wire-wound resistor and is it reliable?

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Messrs. D. R. Chick and **D. P. R. Petrie** (*in reply*): Dr. Boag has mentioned the surprising breakdown of Paschen's law in Van de Graaff generators, with their relatively smooth geometry. In the spherical part of our generator the field falls by a factor of two in passing from the top terminal to the inner surface of the intershield, and by the same factor from the outer surface of the intershield to the tank wall. This amount of field inhomogeneity is apparently enough to account for the failure of Paschen's law and also for the observed difference between the results with positive and negative polarity. We agree that little is known about breakdown at these voltages.

In principle, metal belt 'guide-bars' would seem to be necessary, and other workers have used them, apparently successfully. In our experience with metal bars, the first two or three from the bottom remove much of the charge from the belt by corona. We agree that this may be due to local stress concentration on the bars because of the wide spacing. Insulating bars were found necessary to ensure that all the charge put on the belt reached the top. The bars then act mainly as mechanical guides.

In reply to Dr. Allan, we find that brass, copper and mild steel corrode fairly quickly and should wherever possible be plated or painted. Aluminium and its alloys and cadmium-plated metal are quite unattacked. We have had no trouble from corrosion on the d.c. generator commutator, and very little trouble with tungsten-tipped relay contacts. We have tried a mixture of nitrogen and carbon dioxide, but found that the voltage per-

formance of the generator (without tube) was worse than with pure nitrogen.

The advantage of a wide accelerating tube with high pumping speed is to remove quickly the sudden large burst of gas that occurs whenever the tube breaks down. If this is not removed before the generator voltage has built up again, breakdown occurs again and again, and the system will not recover until the spray current is reduced. Another advantage of the wide tube is to minimize the chance of impact of straggling protons from the beam with the electrodes.

Mr. Moore gave a most interesting account of the Medical Research Council accelerator, and we note particularly his success with diaphragms to waist the tube, and the apparent insensitivity of his tube to oil vapour. We have had success with charging the down-run of the belt by a system of spray combs, similar to Mr. Moore's, but have found that it impairs stability of the generator a little.

Mr. Howard-Flanders has described the beam-focusing system used in the Medical Research Council accelerator, which has the advantage that the focusing is independent of the operating voltage over a wide range, although there must be some sacrifice in maximum voltage attainable because their tube is not uniformly stressed. Our focusing system as described in the paper was very satisfactory for the first tube, for which it was designed. A similar system was incorporated in the second tube without redesign, but did not focus well above 2 MV. In the third tube we are using a system based on Elkind's* principles, which has the merit that the focus setting is relatively independent of ion-source conditions. This has so far been tried up to 3·25 MeV and appears to be satisfactory, giving a spot of 1 or 2 mm diameter 7 ft below the accelerating tube.

In reply to Mr. Fortescue, the conception of maximum field strength is presumably justified provided Paschen's law is valid, and our working pressure is quite close to the point of departure from Paschen's law.

In Figs. 7 and 8 the results were obtained with about 10% Freon in the nitrogen.

The new design of comb was used for spraying charge on to the belt at the bottom, not for stripping. It consisted in surrounding the entire spray bar with an insulated cylindrical metal tube with a longitudinal slot in the wall through which the needles just protruded. Ideally the shield should be held at some potential between spray potential and earth, but in practice it seems satisfactory to leave the shield floating. The result is that the spray voltage required for a given current is reduced to a little more than one-half of the value required for an unshielded comb. The maximum working pressure is still limited to 150 lb/in^2, however, owing to the onset of surface flashing over the belt, which often seems to be initiated by a spark rather than a corona type of discharge from the spray comb.

* ELKIND, M. M.: 'Ion Optics in Long, High Voltage Accelerator Tubes,' *Review of Scientific Instruments*, 1953, **24**, p. 129.

The time lags of breakdown were measured by applying a step-function potential to the gap under test by means of a 100 kV set with a triggering spark-gap in series. The sudden appearance of potential across the test-gap and its disappearance when the gap broke down were followed on a fast oscillograph. With vacuum gaps the time lag was hardly observable, i.e. small compared with 0·1 microsec. With high-pressure gaps the time lag lay in the range 1 microsec to 1 millisec depending on the geometry of the gap and the voltage applied.

Dr. S. E. Hunt and **Messrs. D. P. R. Petrie, K. Firth** and **A. J. Trott** (*in reply*): In reply to Dr. Allan, the presence of X-rays from the top terminal does add to the background observed on a Geiger–Müller tube used for observing the γ-ray yield, but this has not been found to be serious.

A typical background count for a Geiger–Müller tube situated in the target room and shielded by 0·25 in of lead is 12 counts/min. When the machine is operating near full energy, this is increased to about 60 counts/min, but is still small compared with the yield of most resonances investigated.

Dr. Nemet asked about the calibration of the generating voltmeter. This was done initially by applying a comparatively small potential (100 kV) to the intershield from an external supply. This could be measured accurately by an oil-immersed resistor chain and microammeter. To obtain a reasonable reading on the generating voltmeter, its sensitivity was increased a calculable amount by temporarily altering the value of a capacitance in the associated amplifier by a known amount, as described by Fortescue and Hall.*

After the installation of the electrostatic analyser a direct measurement of beam energy, and hence terminal voltage, was available. This made possible a direct check of the calibration by running the generator with the upper half short-circuited; i.e. with the terminal and intershield at the same potential. The calibration of the generating voltmeter has been found to be constant to a fraction of one per cent.

Under normal working conditions, however, the generating voltmeter measures the potential of the intershield. This is not always exactly proportional to the top-terminal potential because of the presence of currents in the tube, and in the gas between the top-terminal intershield and tank, which effectively shunt the stack resistance chain. A variation of few per cent in generating voltmeter indication can therefore be obtained for a given top-terminal potential.

The 50-megohm wire-wound resistor mentioned by Dr. Nemet is made up of commercial 100-kilohm wire-wound units as described in the paper. The stack, complete with anti-corona shields, stands in an earthed metal shield about 6 ft high and about 3 ft in diameter. Repeated measurements of the voltage-divider ratio by the method described in the paper have shown this to be constant to better than 2 parts in 10^4.

* FORTESCUE, R. L., and HALL, P. D.: 'The High-Voltage Electrostatic Generator at the Atomic Energy Research Establishment,' *Proceedings I.E.E.*, Paper No. 829 March, 1949 (**96**, Part 1, p. 77).

621.317.729

Paper No. 1837 M
May 1955

AN ELECTROLYTIC-TANK EQUIPMENT FOR THE DETERMINATION OF ELECTRON TRAJECTORIES, POTENTIAL AND GRADIENT

By D. L. HOLLWAY, B.E.E., M.Eng.Sc., D.Sc.Eng.

The paper was first received 13th September, 1954, and in revised form 25th January, 1955. It was published in May, 1955, and was read before a JOINT MEETING *of the* MEASUREMENT AND CONTROL *and* RADIO AND TELECOMMUNICATION SECTIONS *15th November, 1955.)*

SUMMARY

A description is given of a general-purpose electrolytic-tank equipment of simple design. Provision is made for testing either axially-symmetric or two-dimensional models, and the associated circuits can be arranged by switching (*a*) to measure potentials, (*b*) to mark equipotentials automatically, (*c*) to measure potential gradients, and (*d*) to trace electron trajectories.

As a test of accuracy a family of parabolic trajectories was compared with calculated paths. The errors were found to be less than 1% of the true co-ordinates, indicating that the accuracy is sufficient for most problems found in the design of electrode structures.

(1) INTRODUCTION

In electronic design it is often necessary to devise an electrode system which will constrain charged particles to move in certain desired trajectories. As usually such problems cannot be solved by direct calculation, the best approach is often to combine calculation with a process of trial and error, the time taken to reach the result depending on the speed and accuracy of the method chosen to solve the electron trajectories in trial electrode systems. For this purpose it has been found that an analogue path computer using an electrolytic-tank model has advantages in speed, accuracy and convenience over alternative methods such as graphical construction and the use of rubber-membrane models.

The first trajectory computers were described by Gabor[1] and Langmuir[2] in 1937, but comparatively few appear to have come into use subsequently.[3–9] It is considered that the practical value of these computers is not sufficiently emphasized in texts and deserves wider appreciation.

The present electrolytic tank and path computer differs from those described previously. It is a general-purpose equipment designed to measure potential and potential gradient, to mark equipotentials and to trace electron trajectories.

(2) PRINCIPLE OF OPERATION

The circuit used for potential measurement is shown in a simplified form in Fig. 1(*a*). The potentials of the electrodes in the tank are supplied by the secondary winding 1–2 of transformer T_1 which is also connected to the measuring potentiometer R_1. To measure the potential V at some point in the liquid without drawing current from the measuring probe, the probe is connected to the high-gain phase-reversing amplifier A_1. The output of A_1 is connected to the supply winding, thus completing a negative-feedback loop which operates to bring exactly to earth potential the point in the liquid where the probe rests. If the potentiometer R_1 is so adjusted that a null point is indicated at N_1, the potential V may be read from the setting of R_1.

For the measurement of potential gradient a double probe is used as shown in Fig. 1(*b*). The second probe is connected to the gradient amplifier A_2 having a gain g. The potentiometer R_1

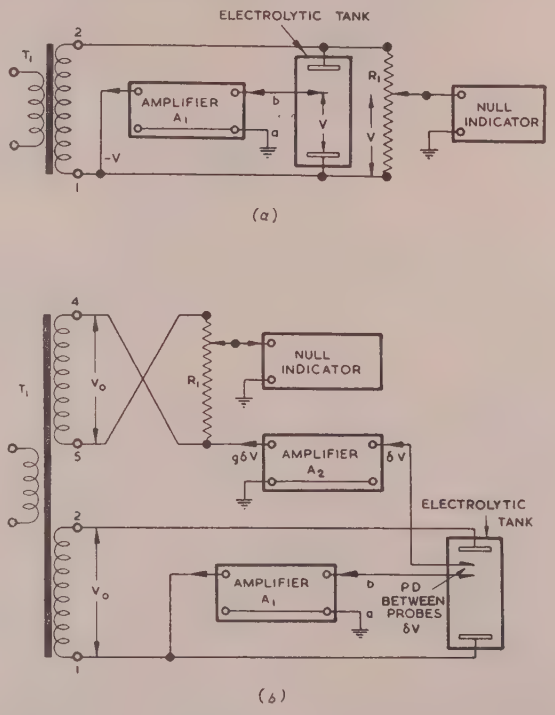


Fig. 1.—Basic circuits used for the measurement of potential and potential gradient.
(*a*) Potential.
(*b*) Potential gradient.

is supplied from a separate winding 4–5 of T_1 exactly equal in voltage to 1–2. If δV is the potential difference between the probes, a null point is observed when the potentiometer R_1 reads $g\delta V$, since the signal at a–b is negligible. Thus the reading of R_1 is a measure of the component of gradient in the direction of the probes.

(3) THE ELECTROLYTIC TANK

The general arrangement of the equipment is shown in Fig. 2. The tank is rectangular (120cm × 60cm, and 30cm deep) and is constructed of 18-gauge brass sheet. Insulating panels, which may be removed when necessary, cover all the internal surfaces. A sheet of Perspex, 94cm × 56cm, serves as a floor for mounting the electrodes and is engraved with a 2cm grid to assist in their positioning. The sheet is supported on a frame designed so that it may be set and locked at any desired level and inclination; thus by setting the floor horizontal, two-dimensional problems may be solved, or by inclining the floor, a wedge of water may be isolated to represent a system having axial symmetry.

The electrodes are of thin copper or brass sheet and are cleaned just before each test by swabbing with a pad of glass

Dr. Hollway is in the Division of Electrotechnology, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia.

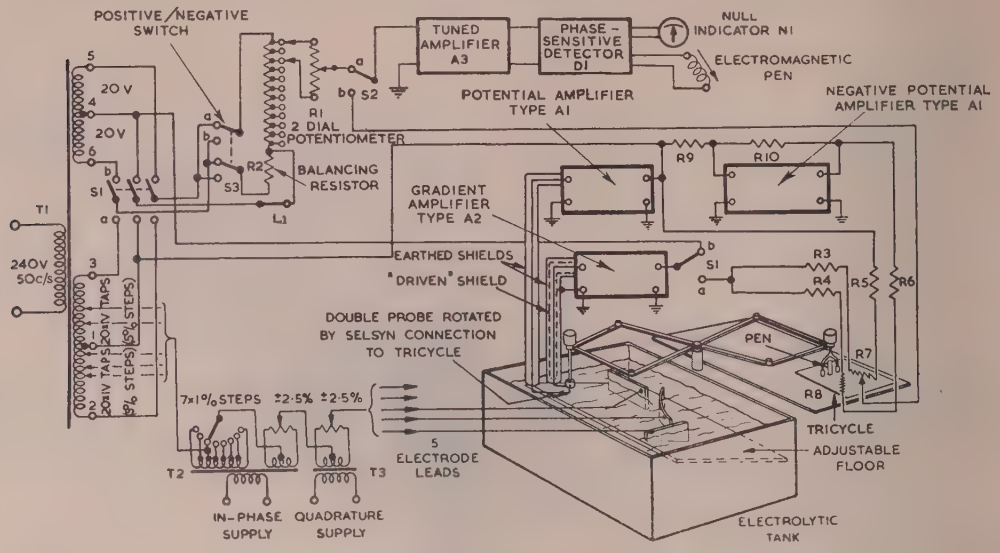


Fig. 2.—Layout of the electrolytic-tank equipment showing the electrode supply transformers and amplifiers.

For different functions the circuits are switched as follows:
Potential measurement: S_1 and S_2 are set in position *a*.
Gradient measurement: S_1 is set to *b*, S_2 to *a*.
Electron trajectory tracing: S_1 is set to *a*, S_2 to *b*.

wool dipped in chromic-sulphuric acid solution, followed by a thorough rinsing. As the electrolyte, plain tap water is preferred, the conductivity being 18×10^{-3} mho/m.

The electrode potentials are derived from the winding 2–1–3 of transformer T_1, which provides ± 20 volts at 50c/s about the centre point 1 and is tapped at one-volt intervals. To avoid confusion with the true potentials of the original electrode system, the taps are marked on a percentage scale between plus and minus 100%. Any number of electrodes may be connected to these taps, and in five leads provision is made to interpolate from tapped windings on T_2. In certain cases it is inconvenient to clean the electrodes and the resulting quadrature components may be eliminated when necessary by adjusting the potentiometers connected to T_3. The impedance of these supplies has been kept low so that the current usually flowing in the liquid does not change an electrode from its nominal potential by more than a few tenths of one per cent. Thus, as the voltage drop at the surface of a clean electrode normally is less than 0·5%, in many problems the electrode potentials may be set without measurement.

Where the highest accuracy is needed, it is often possible to establish virtual electrodes in the liquid itself by displacing the metal electrodes a few millimetres and adjusting their supplies until the required potentials are measured at the correct positions in the liquid. These "liquid electrodes" also have the advantage of not interfering with the movement of the measuring probes.

Beside the tank there is a drawing board slightly larger than the tank and covered with a sheet of plate glass to provide a true surface. A pantograph covers the working areas of both tank and board as shown schematically in Fig. 2. Coupled Selsyns are attached to each end of the pantograph, that at the tank end carrying the double probe.

(3.1) The Measuring Probe

In order to measure gradient accurately the probes must be closely spaced and introduced into the liquid in a way which does not cause appreciable disturbance of the current distribution. In particular, when there is a meniscus near the probes, variations of several per cent of the gradient reading are usual, and it has been found preferable to submerge the probes completely. In designing a submerged probe it is necessary to reduce the volume of insulating material near the tips to the minimum consistent with rigidity. The final form, shown in Fig. 3, is based

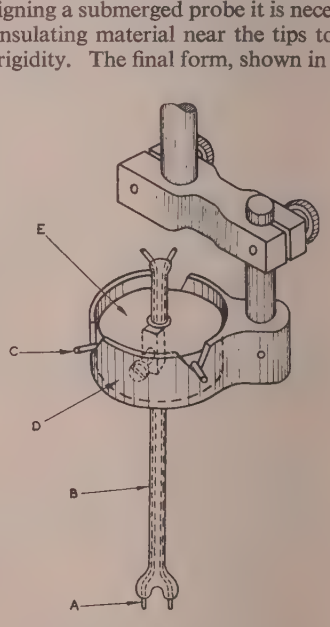


Fig. 3.—Twin-probe construction and mounting.

The probe assembly is arranged to tilt and rise freely from the supporting ring on accidental contact with an electrode.

on these ideas. The stem B is of glass fused about two Dumet wires which project 1 mm as shown at A. The stem is clamped in a brass block E supported by three equally spaced pins C resting horizontally in vee notches in the outer ring D. Thus on accidental contact with an electrode the whole probe is able to tilt or rise freely, subsequently returning to its previous

osition. Provision is made for adjusting the height of the robe and for setting it exactly on the centre line of the Selsyn haft.

In a uniform field the potential difference between the probes alls slightly as the depth of submersion is reduced. For the resent design the gradient reading remains constant as the epth is varied from 40 to 19mm, and decreases by $1{\cdot}3 \times 10^{-3}$ f that reading at 10mm and $3{\cdot}5 \times 10^{-3}$ at 4mm. Normally he probe is submerged as far as the conditions of the problem llow.

(4) POTENTIAL MEASUREMENT AND EQUIPOTENTIAL TRACING

Potentials are measured and equipotentials traced by setting witch S_1 to position *a* in Fig. 2. In this position R_1 is connected etween the centre tap 1 of T_1 and either the positive tap 2 or egative tap 3. R_1 is a two-dial potentiometer of the Kelvin–Varley type, marked in 0·1% steps. All potential and gradient measurements are read from these dials.

The winding 2–1–3 of T_1, and consequently the electrodes and he body of liquid in the tank, are connected to the output of the otential amplifier A_1. The input to A_1 is the potential difference etween one of the measuring probes and earth. The gain of the otential amplifier is made effectively infinite by the addition of a small amount of internal positive feedback. As the transformer windings and electrolyte complete a negative-feedback oop, the output potential is always just sufficient to bring the iquid surrounding the probe exactly to earth potential. Thus when the potentiometer R_1 is adjusted manually so that a null oint is indicated at N_1, the potential difference between the robe and the zero or cathode point 1 of T_1 may be read direct rom the dials of R_1. To measure potentials of the opposite ign, S_3 is switched to position *b*.

Resistor R_2, equal to R_1, balances the loading of R_1 on the pposite winding of T_1 and allows the equality of the two windings o be checked by setting R_1 to zero and observing that N_1 remains nchanged when the link L_1 is opened.

(4.1) The Potential Amplifier

The two amplifiers marked A_1 in Fig. 2 are identical high-gain amplifiers having the circuit shown in Fig. 4. Since the esistance of the probe considered as a generator is high (10^4 to 10^5 ohms) it is desirable that the impedance loading the probe hould exceed 10^8 ohms so that the current drawn will introduce a negligible error. The probe is connected direct to the grid of V_1 and the usual grid resistor is omitted. This is made possible by designing the amplifier A_1 to have appreciable d.c. gain so that he negative-feedback loop including the tank provides a return ath for the grid. This also simplifies the problem of keeping the eedback loop stable at low frequencies.

At the operating frequency of 50c/s the gain of A_1 is made effectively infinite by the addition of the positive-feedback loop ncluding R_7, R_8 and R_9. When the feedback control R_8 is correctly adjusted, as described later, the residual input signal at maximum output is less than 0·5mV. The phase shift ntroduced by the feedback stabilizing capacitor C_1 is corrected at the working frequency by C_4, and the ratio of these capacitors may need adjustment for complete balance. This adjustment to zero of the signal appearing between the grid of V_1 and earth serves two purposes. First, as it is not convenient to place the amplifier directly above the probe, the capacitance of a length of shielded cable appears shunted across the input terminals of A_1. As the potential across the cable is made zero, however, this capacitance does not load the probe and the desired high input mpedance is maintained. Secondly, it ensures that the input to he gradient amplifier does not include an error proportional to

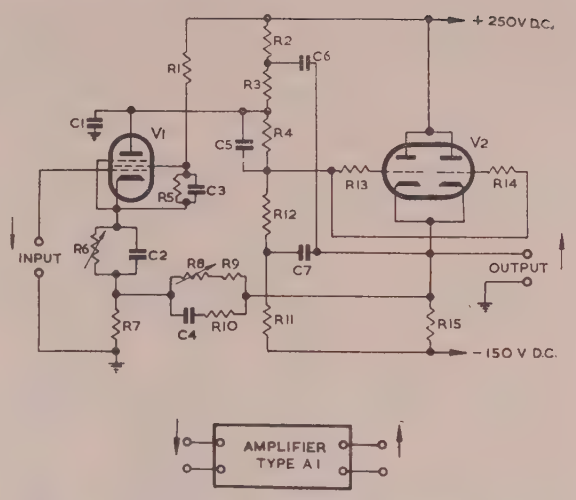


Fig. 4.—Circuit of the potential amplifier A_1.

R1	3·3 MΩ	R9	120 kΩ	C1	500 μμF
R2	100 kΩ	R10	20 kΩ	C2	500 μF
R3	500 kΩ	R11, 12	1·5 MΩ	C3	2 μF
R4	2·5 MΩ	R13, 14	5 kΩ	C4	0·012 μF
R5	1·5 MΩ	R15	10 kΩ	C5	0·05 μF
R6	5–10 kΩ			C6	0·5 μF
R7	25 Ω	V1	VR65	C7	0·1 μF
R8	20 kΩ	V2	12AT7		

the potential. The output impedance of A_1 when connected in circuit is about 0·4 ohm.

(4.2) Equipotential Tracing

For setting electrode potentials or making measurements at a few positions in the model it is sufficiently rapid to balance R_1 while observing the null indicator, but for tracing families of equipotentials a quicker method is desirable and this is provided by the phase-sensitive detector D_1 (Fig. 2) and the electromagnetic pen. The detector is so designed that whenever the in-phase component of the input signal falls below a certain minimum amplitude, an electromagnetically operated pen is brought into contact with the paper. The pen assembly consists of a 1 000-ohm telephone relay linked to a ball pen sliding in guides slightly inclined to the vertical. The inclination and the fact that the pen magnet is pulsed at 50c/s ensure a rolling action at each contact with the paper which renews the ink film. To trace an equipotential, R_1 is first set to the required value. The pen is then moved backwards and forwards over the approximate position and the line is painted in as a succession of dots.

(4.3) The Phase-Sensitive Detector

The full-wave detector shown in Fig. 5 was designed to overcome the need for critical adjustments found in earlier circuits. The input signal is fed to the grid of V_6 from A_3 (Fig. 2), a tuned amplifier having a gain adjustable to a maximum of 100. As the two sections of V_6 have a common cathode resistor, the anodes supply approximately push-pull signals to the cathodes of the switching tubes V_4 and V_5, while the mean anode potentials are stabilized against changes in valve characteristics by negative-feedback loops formed by the resistive dividers R_{19}, R_{22} and R_{20}, R_{23}. The outputs from the switching tubes V_4 and V_5 are compared by the differential amplifier V_3. When the currents in two halves of V_3 balance, the grid of the thyratron tends to be drawn positive by the current flowing through R_2. The thyratron then fires and the pen marks the position of the equipotential. Unbalance in either direction causes the appropriate diode to

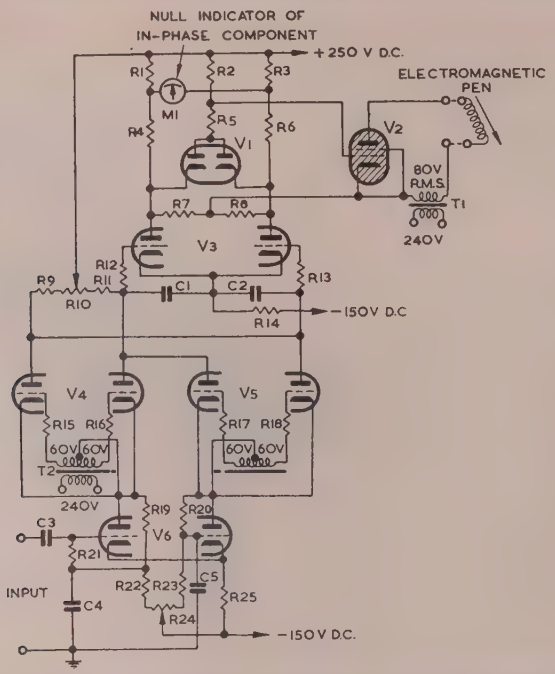


Fig. 5.—Circuit of the phase-sensitive detector and equipotential marker.

R1	15 kΩ	R7, 8	500 kΩ	R15–18	200 kΩ	C1, 2	0·5 μF	M1	500/500 μA
R2	250 kΩ	R9	100 kΩ	R19–20	1·5 MΩ	C3	0·5 μF	V1	6AL5
R3	15 kΩ	R10	20 kΩ	R21	1 MΩ	C4	0·5 μF	V2	2D21
R4	100 kΩ	R11	100 kΩ	R22–23	680 kΩ	C5	0·5 μF	V3	12AT7
R5	25 kΩ	R12, 13	10 kΩ	R24	50 kΩ	T1	Prim. 240 V, Sec. 80 V (r.m.s.)	V4, 5	12AU7
R6	100 kΩ	R14	150 kΩ	R25	15 kΩ	T2	Prim. 240 V, Sec. 2–60/60 V	V6	12AT7

cut off V_2 whenever the potential across R_5 is exceeded. With the values shown, the thyratron is cut off whenever the potential between the two anodes of V_3 reaches 25 volts, corresponding to an in-phase signal at the input terminals of 50 mV (r.m.s.). Thus when A_3 is set to maximum gain the pen marks over an interval of 0·005% of the total potential range, and for normal use it is convenient to reduce the sensitivity by adjusting the gain of A_3.

(5) GRADIENT MEASUREMENT

For the measurement of potential gradient switch S_1 is set to position *b* so that the potential across R_1 is supplied from the winding 6–4–5 of T_1, which is identical with winding 2–1–3. As before, the potential amplifier A_1 operates to bring the liquid surrounding the probe to earth potential, so that the gradient component is proportional to the potential difference between the second, or gradient, probe and earth. This difference potential enters the gradient amplifier A_2 through capacitor C_1 to the grid of V_1 shown in Fig. 6. To maintain the necessary high input impedance, resistor R_8 is adjusted so that no signal appears either between the grid and the shields of the probe cable and input components or across R_1. Thus these impedances do not load the probe, and when R_8 is correct a 50-megohm resistor may be inserted in the probe lead with negligible change in the gradient reading.

The gain of amplifier A_2 is determined by the resistance divider between the output and the cathode of V_1 and earth which forms a negative-feedback loop. Thus R_{10} may be used to set the overall gain in the range 40–60, and this control must be readjusted whenever the double probe is replaced to compensate for variation in the effective distance between the probes.

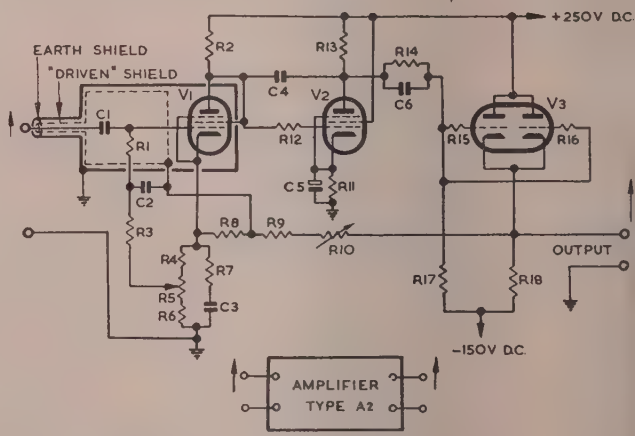


Fig. 6.—Circuit of the gradient amplifier A_2.

R1	1·5 MΩ	R10	20 kΩ (wire wound)	C1, 2	0·5 μF
R2	100 kΩ	R11	120 kΩ	C3	0·035 μF
R3	500 kΩ	R12	5 kΩ	C4	500 μμF
R4	250 Ω (wire wound)	R13	220 kΩ	C5	80 μF
R5	500 Ω (wire wound)	R14	3·3 MΩ	C6	0·1 μF
R6	250 Ω (wire wound)	R15, 16	5 kΩ		
R7	20 kΩ	R17	3·3 MΩ	V1	VR65
R8	25 Ω (wire wound)	R18	10 kΩ	V2	6AM6
R9	40 kΩ (wire wound)			V3	12AT7

Capacitor C_3, in the cathode circuit of V_1, is used to compensate for the phase shift introduced by the feedback stabilizing capacitor C_4. Resistor R_5 is used to set the d.c. level of the output to earth potential.

(5.1) Calibration for Gradient Measurement

Originally, a small test-cell was fitted in one corner of the tank nd used for setting the gain of the gradient amplifier before each roup of tests. It was found that probes of the later design held heir calibration more accurately than measurements could con-eniently be made in a small cell. The cell now used has two arallel plane electrodes 23 cm × 76 cm spaced 38 cm apart and oined by Perspex end-panels and a floor. By careful levelling he gradient can be made uniform to within ±0·1% over a onsiderable part of the area. For calibration the gradient in he cell is determined from the spacing of the equipotentials. Potentiometer R_1 is then set to this value, and the gain adjustment R_{10} (Fig. 6) is varied until a null point is indicated by the detector. The probes are then turned through 180° and the balance is checked with S_3 reversed. The setting of the positive-feedback control R_8 in A_1 (Fig. 4) may be checked also by connecting both electrodes, tied together, to taps 1, 2 and 3 of T_1 in turn. R_8 is then adjusted so that the gradient reading on R_1 is zero at all taps, indicating that the error voltage at the input of A_1 is zero.

As a potential difference of 7 μV between the probes can readily be detected, the electrical sensitivity is always greater than is needed for making the mechanical adjustments, and, in general, the precision of the equipment is limited by mechanical rather than electrical considerations.

(6) ELECTRON TRAJECTORY TRACING

For tracing electron trajectories the magnetic pen is removed and the tricycle attachment, shown in Fig. 7, is fitted to the pantograph. The geometry of the tricycle is shown in Fig. 8. The path is drawn by the pen F, which corresponds in position to the potential probe in the electrolyte. The radius of curvature of the path is determined by the direction of the wheel D, which may be steered about a vertical axis. A moving arm, in the same plane as the wheel D, makes contact with the inner edge of a pair of uniformly wound resistors JA and AK, at C.

The operating principle is similar to Gabor's[1] tangent bridge, and to describe the tracer it is necessary briefly to repeat the geometrical relationship underlying this method. Consider the motion of an electron at a position corresponding to that of the pen F in Fig. 8. The electric field is such that the gradient probe is more positive by an amount δV. If the electron velocity is v, the potential through which it has fallen is V, the radius of the path is r and the effective distance between the probes is w:

$$mv^2/r = 2ev/r = e\delta v/w \quad . \quad . \quad . \quad . \quad . \quad (1)$$

When the equipment is set for electron tracing the amplifiers maintain point K at the potential $-V$, J at $+V$ and A at $g\delta V$. Thus, from the geometry of the tracer,

$$r/d = \mathrm{DB/BC} = \mathrm{DB/BA} = \mathrm{KC/CA} = V/g\delta V \quad . \quad . \quad (2)$$

since the contact arm C is at zero potential and draws no current.

Thus
$$r = dV/g\delta V \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Combining eqns. (1) and (3), the correct radius of curvature will be drawn if the gain of the gradient amplifier g is given by

$$g = d/2w \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Since $d = 25{\cdot}4$cm and w is nominally $2{\cdot}5$mm, g is approximately 50. The exact value is set using the test cell as described previously.

To trace a path, the tricycle is first placed at the desired starting-point of the trajectory, which need not be the cathode, and aligned in the desired direction. The sharp edge of the rear wheel provides an accurate means of setting the initial direction. The contact arm is then moved to the balance point and the null point is followed as the carriage is moved forward. The curvature is not restricted—any radius between zero and infinity may be traced in either direction.

Although it would not be difficult to include servo-motors for automatic balance and carriage drive, the time saved would be small, since usually each path takes less than one minute to trace and the greater part of the time is always spent preparing and setting up the model.

(6.1) Tricycle Construction and Calibration

The general design of the tricycle is shown in Fig. 7. The body consists of a sheet of Perspex 30 cm × 33 cm made rigid by a metal frame. The three wheels are hardened and ground

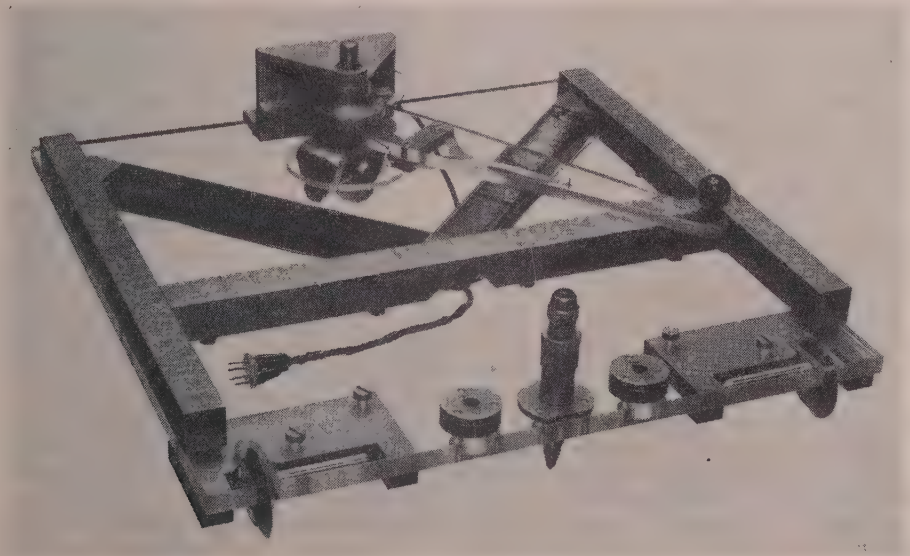


Fig. 7.—Tricycle used for tracing electron trajectories.

integral with their shafts, which run in miniature ball-races. The edges of the wheels are made as sharp as possible without cutting the paper. On each side of the pen and in line with front axles there is a socket in which is fitted the yoke attached to the Selsyn shaft at the end of the pantograph. The potentiometer windings, of 12 kilohms each, are wound on Glyptal-coated brass tubes shaped so that the inner edges (JA and AK in Fig. 8) are raised to allow the moving arm to make only a point contact.

For accuracy, care must be taken in the construction of the tricycle. Fortunately even small errors are readily shown up by simple tests on the drawing board. For example, when the moving arm is set to the zero point on the curvature scale, the trace should be a straight line having a bow less than 0·5mm/metre and bisected throughout by the identation of the rear wheel. Also, when set to either of the zero-radius positions, the trace made by the pen on rotating the tricycle should be a dot and not a small circle or spiral.

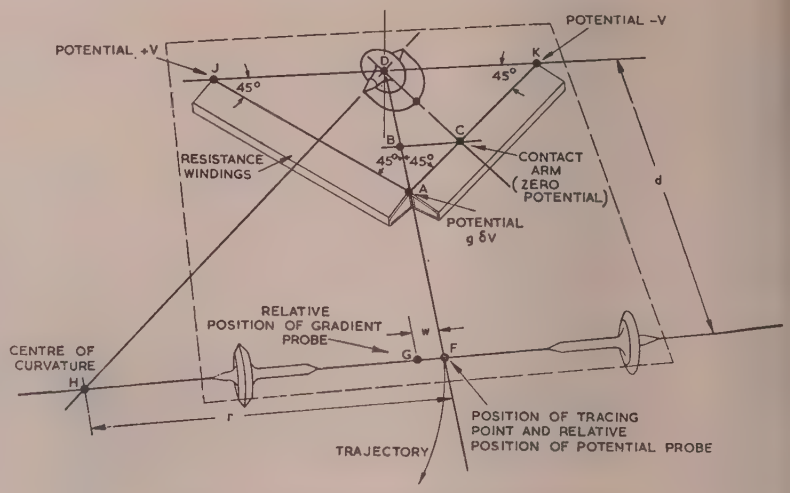


Fig. 8.—Geometry of the electron tracer tricycle.

When the tracer had been adjusted to pass these tests satisfactorily a calibration was made to ensure that the electrical resistance ratio was accurately related to the path curvature. This was done by measuring the radii of circles or arcs drawn with the contact arm at twelve settings along each resistor. In these positions also, the electrical resistance ratios R_{AC}/R_{AK} were measured by a bridge method to eliminate the effect of varying resistance at the contact arm. The error in the resistance ratio α given by

$$\alpha = \frac{1}{(r/d)+1} - \frac{R_{AC}}{R_{AK}} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

was found to increase from zero near the centres of the windings to about 0·01 at the ends. The addition of the small correcting resistors R_3–R_6 in Fig. 2 reduced α below 0·002, which corresponds to a negligible error in curvature.

(7) ACCURACY

The electronic circuits have been designed so that the errors they introduce do not appreciably reduce the accuracy obtainable from the tank-and-probe mechanism. The negative-feedback loops included in amplifiers A_1 and A_2 reduce the effects of changes in valve characteristics. Thus a change of 10% in the internal gain of A_1 would cause an error of 0·002% in the potential reading and a resulting error in gradient of 0·1% of the range of potentiometer R_1. This is equivalent to a curvature which would displace the end of a trajectory 50cm long by

Table 1

Trajectory angle	Error in maximum height	Error in range
	%	%
70°	< − 0·1	− 0·8
60°	< + 0·1	− 0·3
45°	+ 0·3	< + 0·1
30°	− 0·2	− 0·5

0·5mm. A 10% drift in the gain of A_2 would cause an error of 0·1 to 0·25% of the gradient reading. However, all these errors are eliminated by the routine described in Section 5.1.

The tracer has been used to solve many problems over a period of years and the results have been found to be highly repeatable. During this time several measurements of the accuracy of the tracings have been made by drawing families of parabolic paths in the test cell described previously and comparing them with true paths calculated from the trajectory angles and the measured distance to the zero equipotential. An indication of the accuracy of the tracer is given by the results of a typical set of four parabolic paths in which the maximum errors in height and range were as shown in Table 1.

(8) ACKNOWLEDGMENTS

The equipment was constructed to assist the research programme of the Division of Electrotechnology, Commonwealth Scientific and Industrial Research Organization. The author wishes to thank Professor D. M. Myers for his interest and for making laboratory facilities available in the Department of Electrical Engineering, Sydney University.

(9) REFERENCES

(1) GABOR, D.: "Mechanical Tracer for Electron Trajectories," *Nature*, 1937, **139**, p. 373.
(2) LANGMUIR, D. B.: "Automatic Plotting of Electron Trajectories," *ibid.*, 1937, **139**, p. 1066.
(3) LANGMUIR, D. B.: "An Automatic Plotter for Electron Trajectories," *RCA Review*, 1950, **11**, p. 143.
(4) MARVAUD, J.: "Traceur automatique des trajectoires électroniques," *Comptes rendus des séances hebdomaires de l'Académie des Sciences*, 1948, **226**, p. 476.
(5) MARVAUD, J.: "Traceur automatique de trajectoires électroniques et son adaptation à la determination des lignes de courant dans un bassin électrique," *ibid.*, 1952, **234**, p. 45
(6) SANDER, K. F., OATLEY, C. W., and YATES, J. G.: "Plotting Electron Trajectories," *Nature*, 1949, **163**, p. 403.
(7) SANDER, K. F., OATLEY, C. W., and YATES, J. G.: "Factors affecting the Design of an Automatic Electron-Trajectory Tracer," *Proceedings I.E.E.*, Paper No. 1175 R, August 1951 (**99**, Part III, p. 169).
(8) SVALA, G.: Private communication. A brief description of this electron tracer is given by Ekelöf, S.: "Les machines mathématiques en Suède," *Transactions of the Chalmers University of Technology*, Gothenburg, Sweden, 1951 No. 116.
(9) BAKER, B. O.: "Automatic Electron Trajectory Plotting using the Electrolytic Tank Analogue," *British Journal of Applied Physics*, 1954, **5**, p. 191.

[The discussion on the above paper will be found on page 163.]

1.317.329

Paper No. 1838 M
May 1955

A METHOD OF TRACING ELECTRON TRAJECTORIES IN CROSSED ELECTRIC AND MAGNETIC FIELDS

By D. L. HOLLWAY, B.E.E., M.Eng.Sc., D.Sc.Eng.

The paper was first received 13th September, 1954, and in revised form 25th January, 1955. It was published in May, 1955, and was read before a JOINT MEETING *of the* MEASUREMENT AND CONTROL *and* RADIO AND TELECOMMUNICATION SECTIONS *15th November, 1955.)*

SUMMARY

It is shown that Gabor's[1] tangent method of electron-path tracing may be extended to problems which include the determination of paths in the presence of crossed electric and magnetic fields. The magnetic field may be either constant or it may vary with position if the distribution is known. The possible extension to time-varying fields is indicated.

Experimental results are given, including a trochoid compared with the calculated path and typical paths in a split-anode magnetron.

(1) INTRODUCTION

The motion of electrons in crossed electric and magnetic fields is of interest in an increasing number of electronic applications. Recently Alfvén and Romanus[2] have shown that electronic counting and switching tubes may be based upon the properties of beams controlled by a constant magnetic field, and the problem of finding trajectories in certain types of magnetron is very similar.

The usefulness of an electron-path tracer based on Gabor's[1] tangent principle, in purely electric problems, led the author to consider the possibility of extending the analogy to predict electron trajectories in combined electric and magnetic fields. It was found that by a simple modification of the tracer circuit the acceleration due to the magnetic field could be included.

(2) PRINCIPLE OF THE METHOD

(2.1) Path Tracing in an Electric Field

The circuit details and construction of a trajectory tracer operating in purely electric fields have been described previously,[3] and the additional equipment needed for including the effects of magnetic fields will be described in association with this tracer. However, the general principle of the method is applicable to other tracers. No mechanical modifications are needed, since the magnetic field is represented by an electric potential.

In the tracer described, the electron path is drawn by a pen attached to a tricycle, shown in Fig. 1, which moves across a drawing of the electrode system. The tricycle is constrained to follow the correct curvature at all points along the path by steering the rear wheel D about a vertical axis. A moving arm, held in the same plane as the rear wheel, makes contact with one of the linear resistance windings KA, AJ at C. The electrical circuits maintain point K at a potential V corresponding to the electron energy, point J at $-V$ and A at $-g\delta V$, where δV is the potential difference between two probes, one at a position in the tank corresponding to the pen F and the second at a distance w along the radius. When the arm C is set to the zero potential point the radius r is given by[3]

$$d/r = \mathrm{AC/CK} = -g\delta V/V \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the curvature is correct when

$$1/r = -d/2w \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

(2.2) Crossed Electric and Magnetic Fields

When a magnetic field B is superimposed perpendicular to the plane of the path, the equation of motion of an electron becomes

$$mv^2/r = e\partial v/\partial r + Bev \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and since $\delta V/w \simeq \partial V/\partial r$ and $mV^2/2 = eV$,

$$1/r = \delta V/2Vw + B/\sqrt{(2mV/e)} \quad . \quad . \quad . \quad . \quad (4)$$

The field B is considered positive when it is directed downwards towards the plane of Fig. 1.

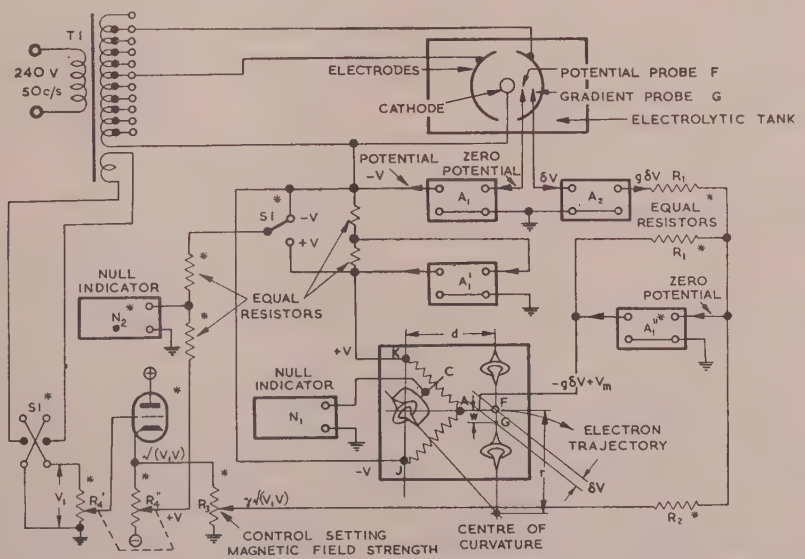


Fig. 1.—Circuit of the computer used for tracing electron paths in combined fields.
Components marked with an asterisk are concerned only with the magnetic field.

Now suppose that at point A a potential V_m is added to the value $-g\delta V$ existing in the electric tracer. The new potential of A is $V_m - g\delta V$, so that eqn. (1), giving the radius of the path traced by the tricycle, becomes

$$d/r = (V_m - g\delta V)/V \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Dr. Hollway is in the Division of Electrotechnology, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia.

When the condition given in eqn. (4) is introduced,

$$V_m = B\sqrt{V}[d\sqrt{(e/2m)}] \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Thus the tricycle will trace correctly the path determined by the combined fields if a potential V_m proportional to both $\sqrt{V}$ and B is added at point A of the potentiometer windings. It is assumed that the field B is either constant or is a known function of position. The term $d\sqrt{(e/2m)}$ is a constant for the equipment.

To test the usefulness of eqn. (6) the tracer used for electric fields was modified as shown in Fig. 1 by the addition of the components shown marked with an asterisk to the original circuit which was essentially unchanged. The operation of the complete tracer will now be described.

(3) EXPERIMENTAL RESULTS

(3.1) The Path Tracer for Combined Fields

A model of a typical electrode system in which electron trajectories are to be found is shown in the electrolytic tank at the top of Fig. 1. The electrode potentials are set by selecting suitable tapping points on the transformer T_1. The two probes marked "potential" and "gradient" are moved by a pantograph and Selsyn to follow, in the model, positions corresponding to F and G on the moving carriage as it traces the electron path across a drawing of the system. A_1, A_1' and A_1'' are identical amplifiers having inherently high gain, with phase reversal and an internal positive-feedback loop, adjusted to make the overall gain infinite. As the cathode electrode in the model is connected to the output of A_1 and the potential probe to its input, A_1 provides an automatic "Wagner earth," bringing the region of liquid surrounding the probe F exactly to earth potential. Therefore, measured with respect to earth, the potential of the cathode and point J is $-V$ and the amplifier A_1' maintains point K at $+V$.

A constant potential $\pm V_1$ is applied from T_1 to the section R_4' of the ganged potentiometer R_4. While tracing a path, R_4 is adjusted to maintain a null indication at N_2. This occurs when the fraction $\sqrt{(V/V_1)}$ of the potential across each section is tapped off by the moving arms. Thus the potential appearing across R_3 is $\sqrt{(V_1V)}$. A fraction γ of this potential is fed from the potentiometer R_3 to an adding circuit, consisting of the amplifier A_1'' and the resistors R_1 and R_2, which adds algebraically the potentials $\sqrt{(V_1V)}$ and $g\delta V$ appearing at the output of the gradient amplifier A_2. Because the output of A_1'' is reversed in phase and the gain is infinite, the potential applied to the point A, which has been called $(V_m - g\delta V)$ in eqn. (5), is found by equating to zero the currents flowing to the input of A_1'':

$$\frac{V_m - g\delta V}{R_1} + \frac{g\delta V}{R_1} + \frac{\gamma\sqrt{(V_1V)}}{R_2} = 0 \quad . \quad . \quad . \quad (7)$$

or

$$V_m = -\gamma(R_1/R_2)\sqrt{(V_1V)} \quad . \quad . \quad . \quad . \quad (8)$$

Thus from eqns. (6) and (8) it is seen that a given value of magnetic field B may be represented by a setting γ of R_4, where

$$B = -\frac{\gamma(R_1/R_2)\sqrt{V_1}}{d\sqrt{(e/2m)}} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

There is an alternative and more direct method of setting the computer to different values of B. In a preliminary test, the tank wall and all the electrodes of the model are set temporarily to any common potential V_s. In this way the model is made field-free and the curvature of every path becomes

$$\frac{1}{r} = \frac{V_m}{dV_s} = +\frac{B}{\sqrt{(2mV_s/e)}} \quad . \quad . \quad . \quad . \quad (10)$$

The arm C of the tricycle is then set to this value on a calibrated curvature scale, and R_3 is adjusted until a null is indicated at N_1. Since δV is zero, this setting corresponds to the correct value of γ.

The time taken to draw each path is from 2 to 10 minutes, depending on its complexity. This could be reduced if necessary by providing servo systems to control N_1 and N_2, but in most problems a longer time is taken in setting up the model.

(3.2) Accuracy

To examine the accuracy of the system as a whole a trochoid was traced and compared with points calculated for the same conditions. For this test a rectangular cell was placed in the tank having two parallel plane copper electrodes 76cm long spaced 38cm apart with Perspex floor and ends. The gradient throughout the working region of the cell was uniform within $\pm 0{\cdot}2\%$. The value of magnetic field used in calculating the points shown in Fig. 2 was set by making a preliminary test with the model "field free" as described in Section 3.1. The electrode potentials were then set and the trochoid was traced from a starting-point near one end of the cell. The accuracy is almost equal to that of paths traced in purely electric fields. After the four convolutions shown, the departure of the trace from the calculated points was 0·5mm in the horizontal direction, the total distance travelled being approximately 150cm. In the vertical direction the error accumulated at the rate of 0·7mm per turn, corresponding to 0·5% of the tracing circle diameter, 14·2cm.

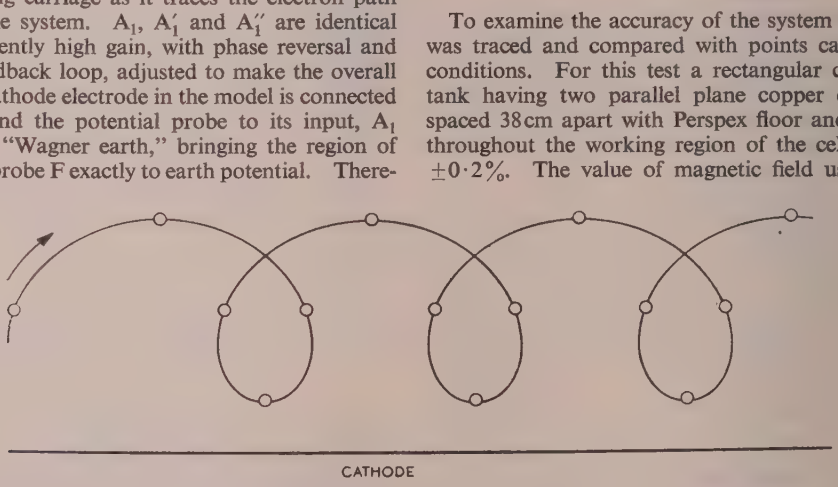


Fig. 2.—A trochoidal electron path traced by the computer compared with calculated points.
The trace has been thickened for reproduction.
—— Traced path.
oooo Calculated points.

(3.3) Paths in a Split-Anode Magnetron

To show the application of the tracer to more complicated problems, a model of the split-anode magnetron was constructed. In work on "negative resistance" magnetrons of this kind, Kilgore[4] photographed electron paths made visible by the admission of a small amount of gas. This procedure was questioned by another writer on the grounds that the presence of sufficient

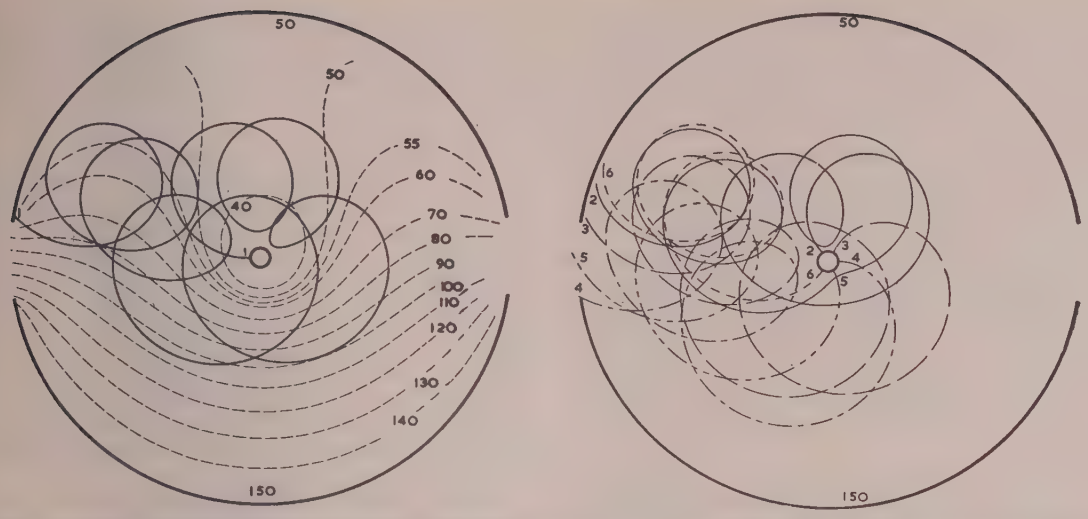


Fig. 3.—Paths of electrons emitted from the cathode of a magnetron, in a magnetic field having 1·5 times the critical value.
It is interesting to note that paths 1, 2 and 3 circle the cathode before settling into a "channel" parallel with the equipotentials and leading to the less-positive anode.

gas to make the paths visible would cause serious distortion. Wallmark,[5] however, investigated Alfvén and Romanus's[2] "trochotron" in this way and found agreement with calculated paths. It was decided to reproduce in the test model the conditions published by Kilgore.

In Fig. 3 is shown a family of trajectories traced from six directions spaced equally around the cathode. It was found that the early part of the path was very sensitive to the direction of emission, and therefore all paths were started radially outwards with an energy of 0·1 volt. There is a general agreement with Kilgore's test, which indicates that little distortion was caused by the gas. In the presence of crossed fields of this kind, electrons follow a trochoidal path with succeeding loops progressing along "channels" parallel with the equipotentials. This tendency is apparent in Fig. 3(*a*). Even after circling around the cathode twice it will be seen that path No. 1 enters the "channel" leading to the less positive electrode.

(4) CONCLUSION

It is shown that an electron-path tracer based on Gabor's tangent method may, by comparatively simple modifications, be adapted to solve electron paths in combined electric and magnetic fields. The magnetic field may be either constant, or a function of position if the distribution is known.

The extension of this method to time-varying fields appears promising. As a potential proportional to $\sqrt{V}$ and to the electron velocity already appears in the circuit, and the distance travelled could be transmitted from the tricycle, the time of travel could be computed automatically. This would allow variations with time of either the electric or magnetic fields, or both, to be fed in while the trace was being drawn.

(5) REFERENCES

(1) Gabor, D.: "Mechanical Tracer for Electric Trajectories," *Nature*, 1937, **139**, p. 373.
(2) Alfvén, H., and Romanus, H.: "Valve with Trochoidal Electronic Motion," *ibid.*, 1947, **160**, p. 614.
(3) Hollway, D. L.: "An Electrolytic Tank Equipment for the Determination of Electron Trajectories, Potential and Gradient" (see page 155).
(4) Kilgore, G. R.: "Magnetron Oscillators for the Generation of Frequencies between 300 and 600 Megacycles," *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 1140.
(5) Wallmark, T.: "Design and Properties of Trochotrons," *Transactions of the Royal Institute of Technology, Stockholm*, **22**, p. 18.

DISCUSSION BEFORE A JOINT MEETING OF THE MEASUREMENT AND CONTROL SECTION AND THE RADIO AND TELECOMMUNICATION SECTION, 15TH NOVEMBER, 1955

Mr. E. R. Hartill: We have been using an electrolytic tank for some time for the measurement of particle accelerating and defocusing potential gradients between various electrode systems. The wedge-tank technique for rotational symmetry as used by the author has been found unsatisfactory, since the depth of the electrolyte tapers to zero at the centre line where the gradients are required with the greatest accuracy. In addition, appreciable errors not mentioned by the author were caused by the differential heating of the shallow electrolyte in this region, since the tap-water used has a temperature coefficient of increase of resistance of approximately 1% per degree centigrade.

For these reasons we have constructed a robust and watertight electrolytic tank made of impregnated wood. It measures 6 ft × 4 ft and is 3 ft deep, and will accommodate a full 180° instead of a wedge model. The two tons of water in the tank are sufficient to maintain a sensibly constant temperature throughout. The plotting table, which is mounted directly above the tank, can be swung into the vertical position to enable a plot to be made in a vertical as well as the more usual horizontal plane. The principle follows that described in a previous paper,* so that complete 3-dimensional electrostatic or magnetic fields can be investigated.

Measurements of the gradients between accelerator electrodes are made using a double-element liquid probe after Sander and

* Diggle, H., and Hartill, E. R.: "Some Applications of the Electrolytic Tank to Engineering Design Problems," *Proceedings I.E.E.*, Paper No. 1627 M, August, 1954 (101, Part II, p. 349).

Yates while the computation of the corresponding particle trajectory is done on a separate computer.*

Meniscus effects at the probes have been reduced by means of a silicone-varnish coating applied as a band above the 2mm depth of immersion. The probe assembly is fitted with a rotating protractor device which enables the axial accelerating gradients, and the much smaller radial focusing or defocusing gradient, to be measured. In this respect we would anticipate difficulties with the author's submerged probe when measuring such small radial gradients on the centre line of a wedge model.

Unlike the author we have found ordinary brass or copper sheet useless as electrode materials, unless their polarization effects are reduced by silver plating. Polarization effects have also been reduced by using a feedback circuit similar in principle to that described in Fig. 1, but using pulse instead of sinusoidal excitation and eliminating the transformer T_1 with its double secondary winding. The above precautions would eliminate the necessity for quadrature correction and the author's artifice of "liquid electrodes."

In a test under somewhat ideal conditions with accurately machined and silver-plated parallel-plate electrodes spaced 10in apart, a gradient of approximately 2% was measured and was repeatable to within $\pm 1\%$. This small error, corresponding to $\pm 0{\cdot}02\%$ on potential measurement, was due to meniscus effects at the electrodes and unavoidable temperature changes in the electrolyte.

Mr. D. H. Davies: At Cambridge we have been working for a number of years on electron trajectory tracing, using a tracer basically similar to that described by Sander, Oatley and Yates,† but using a digital computer instead of a differential analyser. Our experience with this machine causes us to appreciate the accuracy obtained by the author. In view of the relatively simple equipment used, it is extremely good. However, the fact that the balancing of the bridge is manual rather than automatic does raise the question of the extent to which accuracy is limited by the human operator. For how long, for example, can tracing be carried out before errors due to operator fatigue exceed those due to the equipment?

On the question of meniscus effects I should like to mention alternative types of probe set which we have investigated. In one type the probes project through a thin insulating disc at the end of the probe column: with a penetration no greater than 1–2mm, variation of effective probe spacing is of the order of a few tenths of one per cent per millimetre change of penetration. Even better than this is the technique of mounting probes in the floor of a flat-bottom tank and moving the (inverted) model relative to the probes, instead of vice versa.

In addition to trajectory tracing in electric fields we have also carried out some work on motion in crossed electric and magnetic fields and in time-varying fields. The effects of magnetic field normal to the plane of electron motion are simulated by modifying the components of electron acceleration in the computer by amounts proportional to the cross velocity components and to the magnetic field. With this set up, some investigations have been carried out on the electron gun of a carcinotron.

On the problem of tracing in time-varying fields, I suggest that since the fields involved are not conservative the electron velocity cannot be calculated from a knowledge of the probe "potential" alone. Thus the time of travel cannot be computed by such simple means.

Dr. G. Liebmann: Interest in automatic trajectory plotting has increased again during the last few years, and much of the equipment that has been proposed is remarkably complicated. It is therefore refreshing to see a return to the simplified measuring technique used by the author. In particular, I like his way of introducing the Wagner earth. It eliminates many of the complications which are normally inherent in the electrolytic-tank technique.

The author gives the measurement errors for a range of probe depths; but in dealing with axially symmetrical problems the depth of immersion that can be tolerated is probably less than the smallest depth quoted. I should like to know what effect it might have on the accuracy of measurement of the gradient if we were to immerse the probe to this smaller depth. It appears from the paper that this measurement was taken with the probe stationary. In many other techniques, especially where there is fully automatic plotting, the probe is moved. This may give rise to additional errors owing to the meniscus effect and so on. However, in the author's technique, in which the probe is not moved continuously but movement can be interrupted during rebalancing, the effect of a moving probe may not be disturbing.

I find the author's scheme of plotting electron trajectories in crossed electric and magnetic fields very ingenious, particularly because he uses such a very simple computing circuit to introduce the magnetic component. I think this is the first trajectory plotter to use this principle.

We have seen some applications of the author's new technique, and I am anxious to know of further applications. It seems that while the designs of a number of plotting devices have been described, little has been heard about work done with them.

In many problems, of course, the lack of accuracy is a factor which may stand in the way. For instance, in the assessment of aberrations in electron lenses the accuracy required would be too high. There is also the question of space charge. The electrolytic-tank technique does not lend itself to a convenient and sufficiently accurate representation of the space-charge distribution, and for this reason we have devised a scheme using a resistance network for plotting the electron trajectories. It is much easier in a resistance network analogue to represent space charge with sufficient accuracy without the limitations inherent in the electrolytic tank.

In conclusion, I should like to ask how the author compares his electrolytic-tank technique with that using drawing paper for plotting trajectories—a method he introduced several years ago.

Mr. B. J. Mayo: I wonder whether the author has thought of using the method of Musson-Genon in which the charge of the beam can be taken into account by varying the depth of the electrolyte in the tank. In order to make this correction three quantities are needed: the current density, which can be obtained from a knowledge of the electron trajectories, the potential, and the potential gradient. These are the three quantities which this equipment is specially designed to measure. At any rate, there is clearly a great need for even an approximate way of obtaining answers to space-charge problems, particularly since, in a large number of devices (e.g. klystrons, magnetrons and cathode-ray tubes), space charge is a very important factor indeed.

In investigating the image properties of electron lenses one needs to investigate thin pencils of rays. For this, very good accuracy is required, for instance, in obtaining information on curvature of field. The position of the focus of the individual pencils depends on the position of the cross-over of two nearly parallel electron rays. I should like to know whether the author has any results on this.

Dr. D. Gabor: It is a great pleasure to me that someone has

* Crowley-Milling, M. C.: "A Computer for Solving some Problems in Connection with Travelling Wave Particle Accelerators," *Journal of Scientific Instruments*, 1954, p. 100.

† "Factors affecting the Design of an Automatic Electron-Trajectory Tracer," *Proceedings I.E.E.*, Paper No. 1175 R, July, 1952 (99, Part III, p. 169).

made use of the principle of an apparatus I devised almost 20 years ago, and made it work. My own apparatus worked once, but after that it gave nothing but trouble. Having seen it work once I was not very much interested in trouble shooting, and the apparatus was soon discarded.

One of the reasons why we did not continue with this work was that at that time, just before the 1939–45 War, there were no suitable problems. The problems came afterwards, soon after the apparatus was thrown out. Before the war there were only two classes of problem—those for which no trajectory tracing could be accurate enough, such as spherical-aberration determination in the electron microscope, and those which could be solved with a pencil and a piece of paper. Things have changed considerably since then, and the problems which have arisen have revived mechanical trajectory tracing in the hands of Sander, Oatley and Yates and Dr. Hollway.

Although the author uses most of the tricks which I used eighteen years ago, and has added a good many of his own, he does not use one, the wattmetric indicating instrument, which is completely insensitive to currents in the quadrature phase and therefore to polarization. Although my original apparatus did not work again, Dr. C. B. Speedy in my present laboratory has made the wattmetric indicator work, and it has never failed in two years of constant use. With this instrument polarization does not matter, because one measures only the component of the field in phase with a standard vector.

The equipotentials might not, of course, agree exactly with the outlines of the electrodes, but this matters little; one takes what the author has called "liquid electrodes." But compensating the quadrature component, as the author has done, is quite unnecessary if one uses a good wattmetric indicating instrument.

In the measurements carried out in my laboratory tap-water was used, and electrodes coated with Aquadag. The biggest error is the capillary effect at the edges. This can be avoided by using a reasonably deep tank and by measuring a few millimetres below the surface.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Dr. D. L. Hollway (*in reply*): I agree with Mr. Hartill that 180° models are needed in some problems where the axial gradient must be measured to a high accuracy. Similar models have been used in the tank described. For electron tracing, however, the much simpler wedge model is preferable and may be built to a larger scale. A wedge angle of 15° allows a probe submerged 3 mm to approach to within one twentieth of the radius of a 10 in model, the axial electron being undeflected.

In the comparison of different electrode materials, measurements made in tap water at 50 c/s show that brass, copper, silver and nickel have comparable, small surface impedances corresponding to a layer of liquid less than 0·15 mm thick; before cleaning the impedance may be ten times larger. In the circuit described, quadrature components are rejected by the phase-sensitive null detector, and the quadrature supplies in the electrode leads are not normally used. The double secondary winding on T_1 is needed for gradient measurements.

The relative advantage of pulse, high-frequency-sinusoidal and mains-frequency electrode supplies cannot be fully discussed here. Less equipment is needed for mains excitation, and experience shows that the accuracy is high and is limited ultimately by the meniscus and other defects of the liquid model rather than by electrical considerations.

The possibility of errors caused by fatigue, mentioned by Mr. Davies, has not been found serious, possibly because the balancing operation is very simple. A null meter is mounted on the tricycle close to the balance arm, and the operator adjusts his speed to the rate of change of curvature of the path. Furthermore, the periods spent tracing are normally short compared with those spent changing the model. Paths drawn under the same conditions on different days agree closely. Considerations of accuracy alone would not justify the provision of automatic balance and carriage drive, but these additions may be worth while in a tracer intended for continuous use.

I thank Mr. Davies for pointing out the error in the last paragraph referring to time-varying fields. The equipment needed would be more elaborate than that proposed.

In reply to Dr. Liebmann, the output from the gradient probe falls as the depth is decreased, as follows:

Depth, mm . .	4	3	2	1
Reading . . .	0·9965	0·9945	0·988	0·975

These changes are not direct errors in measurement, however, since the gradient-amplifier gain setting is made at the depth to be used in the test.

The tank and tracer were needed originally in the development of an electron-beam decimal counter tube, but the double-probe method of gradient measurement has also proved useful in power-engineering problems. I agree that the accuracy is not sufficient for determining lens aberrations, although useful comparative tests can be made when the curvature is sufficient to allow a model of reasonable radius.

It is considered that many trajectory problems involving space charge can be solved in the electrolytic tank. A method of representing charge clouds, using small numbers of current sources, was suggested by the resemblance between the field about a column of charge and that above a submerged electrode of appropriate dimensions.* When certain rules for the placement, dimensions and current strengths are followed the number of sources is reduced, so that their adjustment is not difficult and the solution may be found after a few trials. This method would be applicable to some of the cases cited by Mr. Mayo.

The method of tracing using ordinary drawing paper, referred to by Dr. Liebmann, is quicker than the electrolytic tank when an equipotential plot of a two-dimensional system is required. For accurate gradient measurements and for electron tracing the tank is more suitable.

I agree with Dr. Gabor on the advantages of a wattmetric indicator. The balance detector of the present tracer performs a similar function and makes the quadrature supplies to the electrodes unnecessary in normal use. In some tests, however, they prove convenient, as, for example, when an oscillograph is used as a null indicator to observe details of the waveform at balance.

* Hollway, D. L.: "The Determination of Electron Trajectories in the Presence of Space Charge," *Australian Journal of Physics*, 1955, 8, p. 74.

621.372.22 : 621.314.2.029.5

Paper No. 1961 R
Mar. 1956

NON-UNIFORM TRANSMISSION LINES AS IMPEDANCE TRANSFORMERS

By J. WILLIS, B.Sc.(Eng.), Associate Member, and N. K. SINHA, Ph.D., B.Sc.(Eng.).

(*The paper was first received 29th July, and in revised form 30th September*, 1955.)

SUMMARY

The equation for the reflection coefficient at one end of a non-uniform transmission line matched at the other end is considered, and an approximate form of solution is suggested which enables such lines to be designed for optimum performance as matching sections. A number of examples are calculated, and the results are compared with lines previously described in literature.

LIST OF SYMBOLS

Γ = Reflection coefficient.
l = Length of non-uniform line.
λ = Free-space wavelength.
x = Distance from the sending end of the non-uniform line.
$Z_0(x)$ = Nominal characteristic impedance of the non-uniform line.
Z_1 = Value of Z_0 for $x = 0$.
Z_2 = Value of Z_0 for $x = l$.
$y = x/l$.
$\mathrm{P}(y) = \frac{d}{dy}(\log_\varepsilon Z_0)$.
d_0 = Inner diameter of the outer conductor of coaxial line.
d_c = Outer diameter of the centre conductor of coaxial line.

(1) INTRODUCTION

The subject of non-uniform transmission lines has attracted renewed interest with the rapid developments in the microwave field during recent years. High-power pulses, used in a number of applications, emphasize the need for a good impedance match over a wide frequency band. Non-uniform transmission lines offer a most attractive solution to this problem since their upper frequency is limited only by the occurrence of higher modes when the transverse dimensions of the line are comparable with the wavelength.

An exact solution of the differential equations for non-uniform lines is possible only in a few special cases. This has led to the development of a number of approximate solutions. Among these, the method developed by Bolinder[3] provides the best approximate solution as regards simplicity and accuracy. Bolinder has applied this method to determine the impedance-matching properties of a number of non-uniform lines with different kinds of taper.

The paper investigates further the theory of non-uniform line impedance transformers with a view to determining the forms of taper for optimum performance under specified conditions.

(2) THEORY

The reflection coefficient at the input end of a loss-free non-uniform line impedance transformer matched at the receiving end is given[3] by

$$\Gamma = \int_0^l \frac{1}{2}\frac{d}{dy}(\log_\varepsilon Z_0) \times \varepsilon^{-j4\pi(l/\lambda)y}dy \quad . \quad . \quad (1)$$

It is evident that Γ will be a function of l/λ, and can be considered as a Fourier transform of the function $\frac{d}{dy}(\log_\varepsilon Z_0) = \mathrm{P}(y)$. It may also be pointed out that $\mathrm{P}(y)$ exists only when y lies between 0 and 1, and that the integral of $\mathrm{P}(y)dy$ between the limits 0 and 1 is always equal to $\log_\varepsilon (Z_2/Z_1)$, where Z_1 and Z_2 are the nominal characteristic impedances of the non-uniform line at its ends.

One of the simplest non-uniform lines is the "exponential line," for which the function $\mathrm{P}(y)$ is constant. In this case,

$$|\Gamma| = \frac{1}{\pi}\log_\varepsilon (Z_2/Z_1)\left|\frac{\sin 2\pi l/\lambda}{4l/\lambda}\right| \quad . \quad . \quad . \quad . \quad (2)$$

The curve showing the relation between $|\Gamma|$ and l/λ has been plotted in Fig. 1, and will be defined as the "reflection pattern."

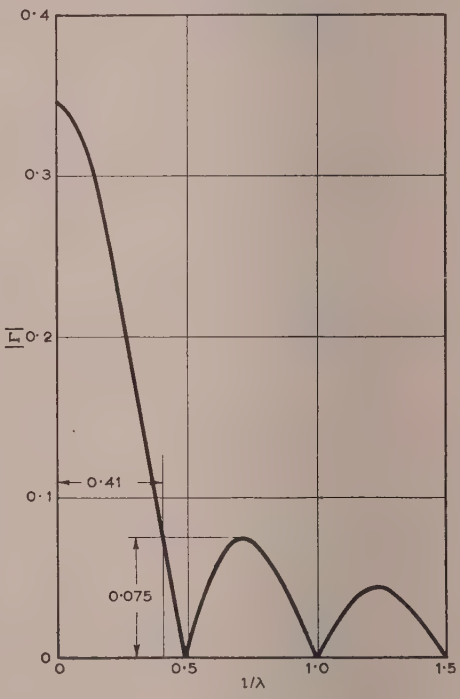


Fig. 1.—Reflection pattern for $P(y)$ = constant.

It will be seen that the reflection pattern has one main lobe and a number of side lobes. The large height of the side lobes indicates that the reflection coefficient is high for certain values of l/λ, i.e. the line is highly frequency-sensitive. For the non-uniform line impedance transformer to give a uniformly good impedance match at all frequencies above a certain specified value, it is essential that the side lobes be as small as possible. It is also desirable that the main lobe be as narrow as possible, in order that the non-uniform line should not become too long.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Willis is in the Faculty of Technology, University of Manchester.
Dr. Sinha was formerly in the Faculty of Technology, University of Manchester, and is now at the Bihar Institute of Technology, Sindri, India.

Hence the optimum non-uniform line can be defined as the one with a reflection pattern having a minimum main-lobe width and vanishingly small side lobes.

It has been shown in an unpublished paper by L. P. Smith[3] that, by increasing the number of continuous derivatives of $P(y)$ at the ends of the non-uniform line, the heights of the side lobes of the reflection pattern can be considerably reduced, but the width of the main lobe is increased. In general, the two requirements of a narrow main lobe and vanishingly small side lobes are contradictory, and a compromise must be achieved in practice. It appears from the Fourier integral energy theorem* that the best results can be obtained when the heights of the successive side lobes decrease very slowly.

Therefore the main problem is to find the particular function for $P(y)$ that will give the optimum relation between $|\Gamma|$ and l/λ. Many attempts have been made in the past to find such a function. The most important of these is the case of the Gaussian function, investigated by Bolinder.[3] As is well known, the Fourier transform of a Gaussian function is also a Gaussian function, so that the resulting reflection pattern will have no side lobes. A difficulty arises, however, in that the Gaussian function vanishes only at infinity, whereas the non-uniform line can only have a finite physical length. The result is a discontinuity in $P(y)$ at the ends of the tapered line, so that side lobes will be produced in the reflection pattern. It has been shown by Bolinder that if the discontinuity is kept small the resulting reflection pattern has side lobes of negligible height.

In the above analysis there is unfortunately no direct method for determining the Fourier transform of the incomplete Gaussian function exactly. A much simpler approach to the problem is the development of a suitable Fourier series for $P(y)$.

It can be shown that if

$$P(y) = k \sin n\pi y$$

$$|\Gamma|_n = \left| \frac{kn}{\pi} \frac{\cos 2\pi l/\lambda}{(4l/\lambda)^2 - n^2} \right| \quad . \quad . \quad . \quad . \quad (3)$$

where n is an odd integer and k is a constant.

(Because $\int_0^1 \sin n\pi y dy = 0$ for even values of n, even values need not be considered, as these do not give an impedance transformation.)

From the above equation it is evident that it is possible to use a Fourier series for the function $P(y)$ containing only odd harmonics of $\sin \pi y$. The coefficients of the various components have to be adjusted in such a way as to reduce the height of the side lobes to the desired value. The advantage of this method lies in the fact that the reflection coefficients due to the various harmonic component can be exactly calculated. Moreover, only a small number of harmonic components will be needed in practice, because $|\Gamma|_n$ is maximum when $4l/\lambda$ is equal to n, this peak value being the same for all values of n. As, in general, for increasing values of l/λ the value of $|\Gamma|$ is required to be smaller, it follows that the coefficient for the nth harmonic must be made progressively smaller as n is increased.

Instead of using odd harmonics of $\sin \pi y$ for $P(y)$, it is also possible to use a constant term and even harmonics of $\cos \pi y$. It can be shown that if

$$P(y) = k \cos n\pi y$$

$$|\Gamma|_n = \frac{k}{\pi} \left| \frac{4l/\lambda \sin 2\pi l/\lambda}{(4l/\lambda)^2 - n^2} \right| \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where n is zero or an even integer.

* cf. Appendix 9.1.

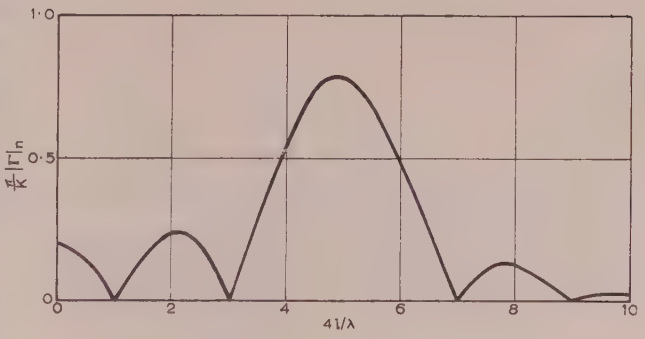


Fig. 2.—$\frac{\pi}{k}|\Gamma|_n$, for $n = 5$.

In this case also, only a small number of harmonics will be needed in practice, from similar considerations as in the previous case.

In either series the constant k will be determined from the condition that

$$\int_0^1 P(y)dy = \log_e \frac{Z_2}{Z_1} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Hence if

$$P(y) = k(\sin \pi y + a_3 \sin 3\pi y + a_5 \sin 5\pi y + \ldots)$$

then

$$k = \frac{\pi}{2} \frac{\log_e (Z_2/Z_1)}{1 + \frac{1}{3}a_3 + \frac{1}{5}a_5 + \ldots} \quad . \quad . \quad . \quad . \quad (6)$$

and if

$$P(y) = k(1 + a_2 \cos 2\pi y + a_4 \cos 4\pi y + \ldots)$$

then

$$k = \log_e (Z_2/Z_1) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The latter series is more suitable when a small width of the main lobe is required, because with the first term only, the exponential line is obtained, the reflection pattern for which has a main-lobe width of $l/\lambda = \frac{1}{2}$. On the other hand, with the first term only, the former series gives a main-lobe width of $l/\lambda = \frac{3}{4}$.

(3) APPLICATION OF THEORY TO SELECTED CASES

A number of distributions of $P(y)$ are obtained which give reflection patterns having a main lobe of given width and side lobes of minimum and substantially constant height.

(3.1) Optimum Tapered Line $\frac{1}{2}\lambda$ Long

In this case, since the width of the main lobe is required to be small, the combination of a constant and $\cos 2\pi y$ for $P(y)$ is most suitable, and very little improvement can be obtained by the addition of small amounts of higher harmonics. By trial a good result is obtained when

$$P(y) = (1 - 0{\cdot}2405 \cos 2\pi y) \log_e \frac{Z_2}{Z_1} \quad . \quad . \quad (8)$$

Hence

$$|\Gamma| = \frac{1}{2} \log_e \frac{Z_2}{Z_1} \int_0^1 (1 - 0{\cdot}2405 \cos 2\pi y) \varepsilon^{-j4\pi(l/\lambda)y} dy$$

$$= |\Gamma_0 - 0{\cdot}2405\Gamma_2| \quad . \quad (9)$$

where

$$\Gamma_n = \left(\frac{1}{\pi} \log_e \frac{Z_2}{Z_1}\right) \frac{4l/\lambda |\sin 2\pi l/\lambda|}{(4l/\lambda)^2 - n^2} \quad . \quad . \quad . \quad (10)$$

Table 1 gives the values of $|\Gamma|$ calculated for an impedance ratio of two (i.e. $Z_2/Z_1 = 2$):

Table 1

$4l/\lambda$	1	2	3	5	7	9
Γ_0	0·2206	0	0·0735	0·0441	0·0315	0·0245
$0·2405\Gamma_2$	−0·0177	0·0417	0·0318	0·0126	0·0082	0·0062
$\lvert\Gamma\rvert$	0·2383	0·0417	0·0417	0·0315	0·0233	0·0183

It is seen (Fig. 3) that a maximum voltage standing-wave ratio (v.s.w.r.) of 1·09 will be obtained for $l/\lambda \geqslant \frac{1}{2}$.

It may be pointed out that the semi-circular distribution of $P(y)$ has been recommended as the most suitable where reflections of the order of 5% can be tolerated.[3] The example considered above

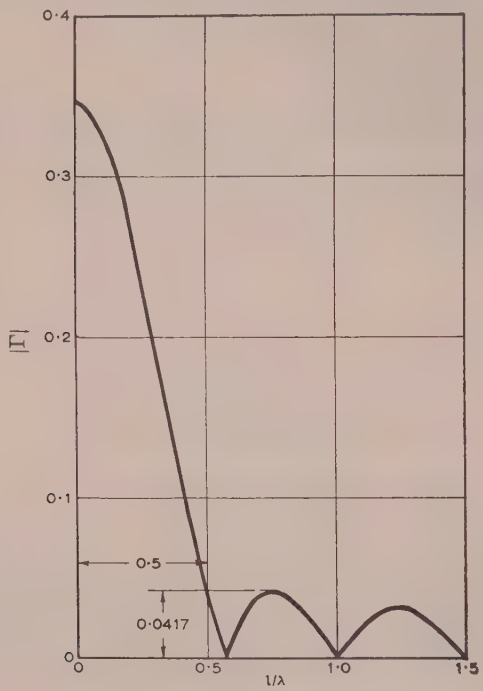


Fig. 3.—Reflection pattern for $P(y) \propto (1 - 0·2405 \cos 2\pi y)$.

reduces the length required by 6% (from 0·53λ to 0·50λ) and reduces the maximum reflection coefficient by about 10% (from 0·046 to 0·0417).

(3.2) Optimum Tapered Line $\frac{3}{4}\lambda$ Long

Here the following distribution gives a good result.

$$P(y) = (1 - 0·636 \cos 2\pi y) \log_e (Z_2/Z_1) \quad . \quad . \quad (11)$$

Hence

$$|\Gamma| = |\Gamma_0 - 0·636\Gamma_2| \quad . \quad . \quad . \quad . \quad (12)$$

Table 2 gives the values of $|\Gamma|$ for an impedance ratio of two:

$$P(y) = (1 - 0·889 \cos 2\pi y + 0·0112 \cos 4\pi y) \log_e (Z_2/Z_1) \quad (13)$$

Hence

$$|\Gamma| = |\Gamma_0 - 0·889\Gamma_2 + 0·0112\Gamma_4| \quad . \quad . \quad . \quad (14)$$

Table 3 gives the values of $|\Gamma|$ for an impedance ratio of two.

It is seen from Fig. 5 that the maximum height of the side lobes is less than 0·002, so that a v.s.w.r. of less than 1·004

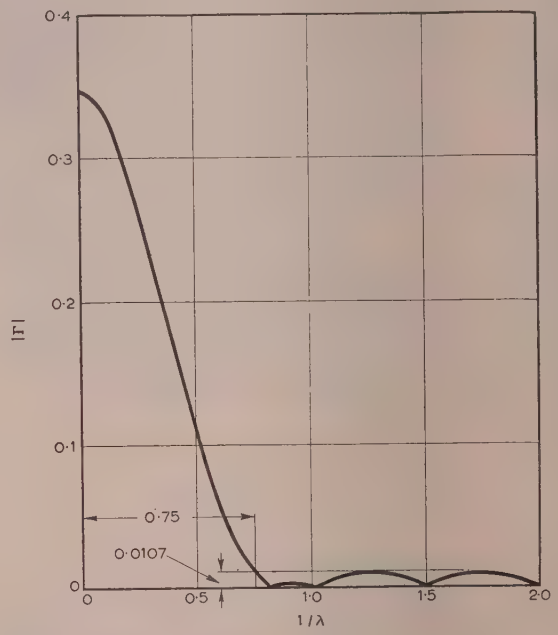


Fig. 4.—Reflection pattern for $P(y) \propto (1 - 0·636 \cos 2\pi y)$.

will be obtained for all wavelengths smaller than the length of the tapered line with an impedance ratio of 2·0. With an impedance ratio of 5·0, a maximum v.s.w.r. of less than 1·01 will be obtained, and this will meet most requirements.

(3.4) Optimum Tapered Line $1\frac{1}{2}\lambda$ Long

Where $P(y) = \sin^3 \pi y$, the reflection pattern has a main-lobe width of $l/\lambda = 1·25$, and has very rapidly decreasing side lobes.[3] This line can be modified to give a main-lobe width of $l/\lambda = 1·5$, and if the modification is made in such a way as to reduce the heights of the first few side lobes, a line of extremely high per-

Table 2

$4l/\lambda$	1	2	3	5	7	9	11
$\lvert\Gamma\rvert$	0·2673	0·1102	0·0107	0·0107	0·0097	0·0081	0·0069

It is seen that this line will give (Fig. 4) a maximum v.s.w.r. of less than 1·022 for $l/\lambda \geqslant \frac{3}{4}$.

(3.3) Optimum Tapered Line One Wavelength Long

In this case, the following distribution was found most suitable:

formance can be obtained. This can be done by reducing the third-harmonic content of the distribution so that the term containing Γ'_3 becomes more nearly equal to Γ'_1 for $4l/\lambda$ greater than or equal to six, or by adding a fifth-harmonic term to give a term containing Γ'_5 more nearly equal to $\Gamma'_3 - \Gamma'_1$; the latter

Table 3

$4l/\lambda$	1	2	3	4	4·65	5
$\lvert\Gamma\rvert$	0·286	0·154	0·045	0·0019	0·00196	0·0012
$4l/\lambda$	7·7	8·95	11	13	15	
$\lvert\Gamma\rvert$	0·00156	0·00192	0·00188	0·00173	0·00157	

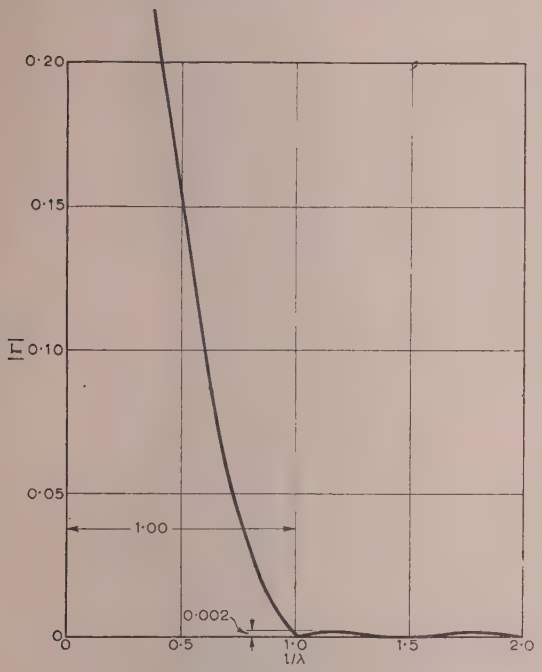


Fig. 5.—Reflection pattern for
$P(y) \propto (1 - 0{\cdot}889 \cos 2\pi y + 0{\cdot}0112 \cos 4\pi y)$.

method gives better results. The following distribution was found to give the least reflection:

$$P(y) = k\left(\sin \pi y - \frac{1}{3}\sin 3\pi y + \frac{1}{54}\sin 5\pi y\right) \quad . \quad (15)$$

where
$$k = \frac{\frac{1}{2}\log_e (Z_2/Z_1)}{1 - \frac{1}{9} + \frac{1}{270}}$$

Hence
$$|\Gamma| = k\left|\left(\Gamma'_1 - \Gamma'_3 + \frac{1}{10{\cdot}8}\Gamma'_5\right)\right| \quad . \quad . \quad . \quad (16)$$

where
$$\Gamma'_n = \frac{|\cos 2\pi l/\lambda|}{(4l/\lambda)^2 - n^2} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Table 4 gives the values of $|\Gamma|$ for an impedance ratio of two:

Table 4

$4l/\lambda$	1	2	3	4	5	6
$\lvert\Gamma\rvert$	0·305	0·205	0·1017	0·03358	0·00565	0·000019
$4l/\lambda$	6·35	8	10	12	14	
$\lvert\Gamma\rvert$	0·000149	0·000025	0·000135	0·000141	0·000126	

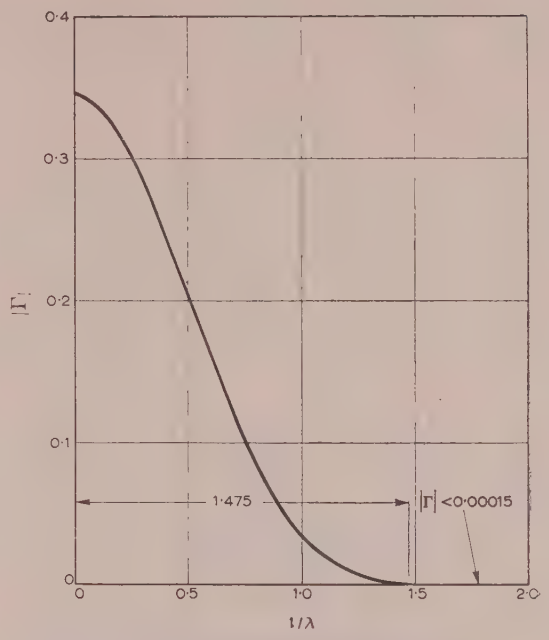


Fig. 6.—Reflection pattern for
$P(y) \propto (\sin \pi y - \frac{1}{3}\sin 3\pi y + \frac{1}{54}\sin 5\pi y)$.

It is seen from Fig. 6 that if $l/\lambda \geqslant 1{\cdot}5$, the maximum value of the reflection coefficient is less than 0·00015, corresponding to a maximum v.s.w.r. of 1·0003. For all practical purposes, this can be considered as being free from reflection.

It is possible to improve upon the above distribution by the inclusion of a very small seventh-harmonic component, but the improvement is very small, and as will be shown later, it cannot be realized in practice.

(4) EVALUATION OF RESULTS

The properties of different types of tapered line have been compared in Fig. 7. In the Figure, r represents the maximum height of the side lobes and w represents the minimum value of l/λ required to obtain a reflection coefficient equal to or less than r. The values of r and w have been plotted for the authors' lines, and a curve has been drawn through them. In the same Figure the values of r and w are given for tapered lines considered by other authors. The improvement conferred by the optimization process is thus illustrated.

(5) PROFILE OF THE CENTRE CONDUCTOR

A convenient form of a tapered coaxial line is one having an outer conductor of constant inside diameter d_0 and an inner conductor of variable diameter d_c shaped to give the desired distribution of $P(y)$. For any given distribution of $P(y)$ the characteristic impedance is calculated for different values of y, and the ratio d_0/d_c determined for each value of Z_0 from the relation

$$Z_0 = \frac{60}{\sqrt{\varepsilon}}\log_e \frac{d_0}{d_c} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The lower frequency, f_1, at which the reflection coefficient of the tapered section is acceptable depends upon the length of the section, which may be 0·5–1·5 times the transverse dimensions of the line, which must be small compared with the wavelength

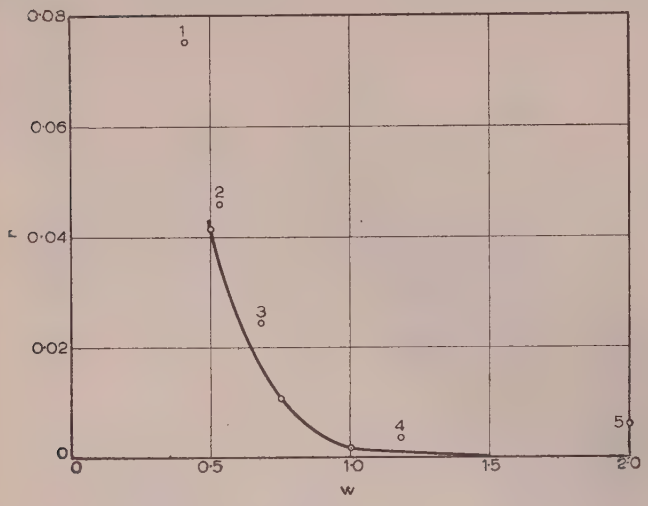


Fig. 7.—Comparison between different types of non-uniform lines.

1. Exponential line (Ref. 3).
2. $P(y) \propto$ semicircle (Ref. 3).
3. $P(y) \propto \sin \pi y$ (Ref. 3).
4. $P(y) \propto \sin^3 \pi y$ (Ref. 3).
5. $P(y) \propto \operatorname{sech}^2 [k(y - \frac{1}{2})]$ (Ref. 11).

In general, the required upper frequency f_2 is so high that d_0 must be made small. For example, for $f_2 = 3\,000$ Mc/s, d_0 should be of the order of 1 in for an air-dielectric line; for such a line, $d_c = 0{\cdot}4346$ in for $Z_0 = 50$ ohms, and $d_c = 1{\cdot}1889$ in for $Z_0 = 100$ ohms.

The achievement of a given accuracy in $P(y)$ is limited by the precision to which the centre conductor can be machined. In normal circumstances it is not possible to achieve a diametral accuracy of better than 0·000 5 in. The percentage error in $P(y)$ resulting from this increases as f_2 increases and d_0 and d_c are decreased.

Small errors in d_c, while causing proportionate variations in $P(y)$, may cause considerable changes in the reflection pattern. For example, comparison between the $(\sin \pi y - \frac{1}{3} \sin 3\pi y)$ distribution and the optimized distribution for $l/\lambda \geqslant 1{\cdot}5$ shows that the latter differs from the former by the inclusion of about 2% of the fifth-harmonic term. The reflection patterns for the two cases differ appreciably; the former has the maximum height of the side lobes equal to 0·003 7 as compared to 0·000 15 in the latter. For the sake of comparison, the two reflection patterns have been plotted together in Figs. 8 and 9 respectively. In both these Figures, curve (*a*) has been drawn for the optimum line.

It will be seen from Fig. 8 that the two distributions differ very slightly, with the result that the profiles of the centre conductors in the two cases also differ very slightly. From Fig. 9 it is evident that the two reflection patterns differ considerably, in the width of their main lobes as well as in heights of their side lobes.

The profile of the centre conductor was calculated for a number of tapered air-dielectric coaxial lines transforming from 50 ohms to 100 ohms and having an outer conductor of inner diameter 1 in. A comparison between the profiles obtained for the distributions shown in Fig. 8 showed a maximum difference of 0·001 1 in, which is just within machining limits. It may be noted that this small change in the profile is due to the addition of a small fifth-harmonic component of magnitude only 1/54 of the fundamental. It follows, therefore, that a smaller change in the magnitude of this component, or the introduction of much smaller amounts of the seventh of the other higher-order har-

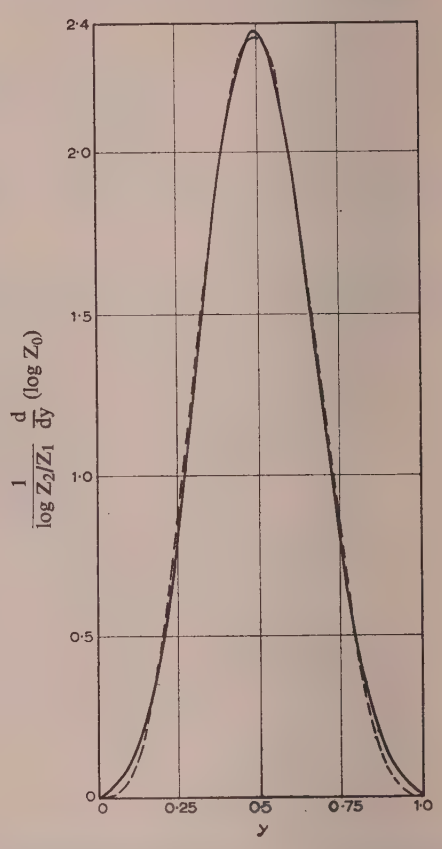


Fig. 8.—Comparison between two distributions of $P(y)$.

(*a*) —— Optimum line 1·5λ long.
(*b*) - - - - $P(y) \propto \sin^3 \pi y$.

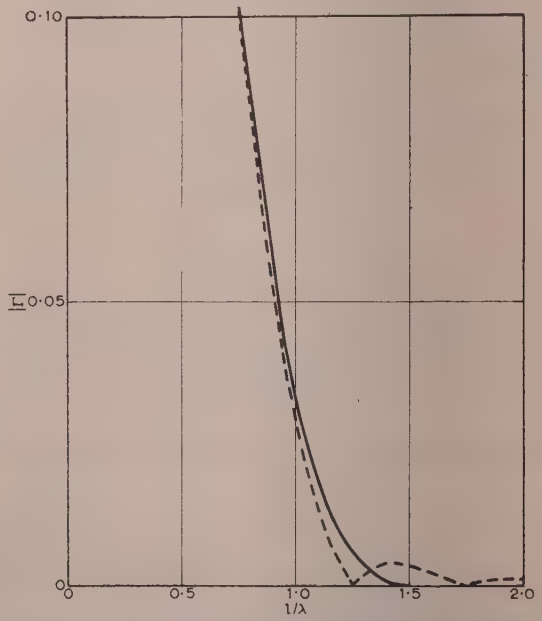


Fig. 9.—Comparison between two reflection patterns.

(*a*) —— Optimum line 1·5λ long.
(*b*) - - - - $P(y) \propto \sin^3 \pi y$.

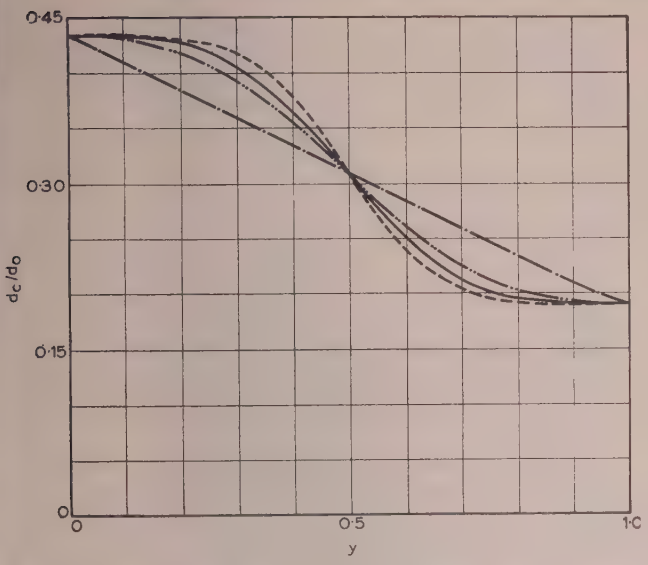


Fig. 10.—Profile of the centre conductor for different types of tapered lines.

(*a*) - - - - - Profile for $P(y) \propto \sin^5 \pi y$.
(*b*) ——— Profile for optimum line 1·5λ long.
(*c*) —·—·— Profile for optimum line λ long.
(*d*) —·— Profile for exponential line.

monics, will represent changes in the profile of the centre conductor less than the machining limits, although the change in the reflection pattern may be appreciable.

Four profiles of the centre conductors have been plotted in Fig. 10. Curve (*a*) gives the profile for the line which has $P(y) \propto \sin^5 \pi y$ (i.e. $\log_\varepsilon Z_0$ has five continuous derivatives[3]), curve (*b*) gives the profile for the optimized distribution for $l/\lambda \geqslant 1{\cdot}5$, curve (*c*) for the optimized distribution for $l/\lambda \geqslant 1$, and curve (*d*) for the exponential line. It will be seen that by increasing the number of continuous derivatives of $\log_\varepsilon Z_0$ at the ends of the tapered line, the change in the ratio d_c/d_0 becomes more gradual near the ends and tends to be concentrated near the centre. This tends to reduce the height of the side lobes in the reflection pattern, but also tends to increase the width of the main lobe. The best profile is represented by curve (*b*). If, as in curve (*a*), the change in diameter is made more gradual near the ends than in curve (*b*), the width of the main lobe of the reflection pattern is increased, since in effect the length of taper is reduced; and at the same time the height of the first few side lobes is also increased. On the other hand, introducing the change in diameter less gradually than in curve (*b*) reduces the width of the main lobe of the reflection pattern, but again increases the height of the side lobes. When higher reflections can be allowed, curves of the latter type are preferable, as they reduce the length of taper required for a given value of the lowest frequency to be transmitted along the line.

(6) CONCLUSIONS

The use of a Fourier series for synthesizing $P(y)$ has been illustrated, and the performance of the optimized lines obtained has been compared with the performance of lines obtained by other authors. Fig. 7 shows that the optimized lines represent a distinct improvement over the lines described previously. The process of optimization can be carried to any desired length, but the accuracy to which the resulting tapered line can be machined sets a limit to its practical realization.

(7) ACKNOWLEDGMENTS

The authors wish to acknowledge the help derived from discussions with many colleagues. Special thanks are due to Prof. E. Bradshaw for his advice and constant encouragement.

(8) REFERENCES

(1) Arnold, J. W., and Taylor, R. C.: "Linearly Tapered Loaded Transmission Lines," *Proceedings of the Institute of Radio Engineers*, 1932, **20**, p. 1811.
(2) Ballantine, S.: "Non-Uniform Lumped Electric Lines," *Journal of the Franklin Institute*, 1927, **203**, p. 561.
(3) Bolinder, F.: "Fourier Transforms in the Theory of Inhomogeneous Transmission Lines," *Transactions of the Royal Institute of Technology, Stockholm*, 1951, **48**, p. 1.
(4) Burkhardtmair, W.: "Widerstand Transformation mit Leitungen," *Funk und Ton*, March, 1949, p. 151, and April, 1949, p. 202.
(5) Burrows, C. R.: "The Exponential Transmission Line," *Bell System Technical Journal*, 1938, **17**, p. 555.
(6) Carson, J. R.: "Propagation of Periodic Currents over Non-uniform Lines," *Electrician*, 4th March, 1921, p. 272.
(7) Clemens, G. J.: "A Tapered Line Termination at Microwaves," *Quarterly of Applied Mathematics*, 1950, **7**, p. 425.
(8) Gent, A. W., and Wallis, P. J.: "Impedance Matching by Tapered Transmission Lines," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 559.
(9) Pierce, J. R. "A Note on the Transmission Line Equation in Terms of Impedance," *Bell System Technical Journal*, 1943, **22**, p. 263.
(10) Rurhmann, A. W.: "Die Energieausbreitung auf Leitungen mit Exponentiell Veranderlichem Wellenwiderstand," *Hoch Frequenz*, 1941, **58**, p. 61.
(11) Scott, H. J.: "The Hyperbolic Transmission Line as a Matching Section," *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1654.
(12) Slater, J. C.: "Microwave Transmission" (McGraw-Hill, New York, 1942).
(13) Starr, A. T.: "The Non-uniform Transmission Line," *Proceedings of the Institute of Radio Engineers*, 1932, **20**, p. 1052.
(14) Wagner, K. W.: "Die theorie Unglichformiger Leitungen," *Archiv für Elektrotechnik*, 1942, **36**, p. 69.
(15) Walker, L. R., and Wax, N.: "Non-uniform Transmission Lines and Reflection Coefficient," *Journal of Applied Physics*, 1946, **17**, p. 1043.
(16) Wheeler, H. A.: "Transmission Lines with Exponential Taper," *Proceedings of the Institute of Radio Engineers*, 1939, **27**, p. 65.
(17) Gill, S.: "Process for Step-by-Step Integration of Differential Equations in Automatic Digital Computing Machine," *Proceedings of the Cambridge Philosophical Society*, 1951, **47**, p. 96.
(18) Van der Maas, G. J.: "A Simplified Calculation for Dolph–Tchebycheff Arrays," *Journal of Applied Physics*, 1954, **25**, p. 121.

(9) APPENDICES

(9.1) Application of the Fourier Integral Energy Theorem

The Fourier integral energy theorem states that

$$\int_0^\infty [\dot{S}(\omega)]^2 d\omega = \pi \int_{-\infty}^{+\infty} [G(t)]^2 dt \quad . \quad . \quad . \quad (19)$$

where

$$S(\omega) = \int_{-\infty}^{+\infty} G(t)\varepsilon^{-j\omega t} dt \quad . \quad . \quad . \quad . \quad (20)$$

Applying this theorem to non-uniform lines,

$$\int_0^\infty |\Gamma|^2 d(l/\lambda) = \frac{1}{16}\int_0^1 [P(y)]^2 dy \quad . \quad . \quad . \quad (21)$$

Hence it follows that in the case of two non-uniform lines having $\int_0^1 [P(y)]^2 dy$ equal, and also the areas $\int |\Gamma|^2 d(l/\lambda)$ equal for the main lobe of the reflection pattern, the maximum height of the side lobes will be smaller in the particular case when the heights of the successive side lobes decrease less rapidly.

(9.2) Accuracy of the Approximation

The correct differential equation for the reflection coefficient at the input end of a non-uniform line matched at the receiving end[3] is

$$\frac{d\Gamma}{dy} + (1 - \Gamma^2)\frac{1}{2}\frac{d}{dy}(\log_\varepsilon Z_0) - j4\pi(l/\lambda)\Gamma = 0 \quad . \quad (22)$$

Eqn. (1) was obtained from eqn. (22) by replacing $(1 - \Gamma^2)$ by unity. This approximation simplifies the non-linear differential eqn. (22) to one which can be easily solved, and is justified if Γ is small compared to unity.

It is of interest to investigate the error introduced by the approximation. This may be readily done by numerical solution of eqn. (22) in some particular case. For this purpose, the real and imaginary parts of Γ must be separated.

Let $$\Gamma = r + js$$

Hence $$\frac{d\Gamma}{dy} = \frac{dr}{dy} + j\frac{ds}{dy}$$

and $$1 - \Gamma^2 = (1 - r^2 + s^2) - j2rs$$

Substituting these values in eqn. (22) and separating the real and imaginary parts,

$$\frac{dr}{dy} + (1 - r^2 + s^2)\frac{1}{2}\frac{d}{dy}(\log_\varepsilon Z_0) + 4\pi(l/\lambda)s = 0 \quad . \quad (23)$$

and $$\frac{ds}{dy} - 2rs\frac{1}{2}\frac{d}{dy}(\log_\varepsilon Z_0) + 4\pi(l/\lambda)r = 0 \quad . \quad . \quad (24)$$

These two simultaneous differential equations can be integrated by using Gill's modification to the Runge–Kutta method.[17] With this method solutions were obtained on the automatic digital computer at the University of Manchester, for the optimum line defined by eqn. (15) for an impedance transformation ratio of two. The values obtained are given in Table 5, together with the values obtained by the approximate method:

Table 5

$4l/\lambda$	1	2	3	4	5
$\|\Gamma\|$ by approximate method	0·305	0·205	0·102	0·034	0·0057
$\|\Gamma\|$ by the computer	0·297	0·205	0·104	0·035	0·0060

$4l/\lambda$	6	7	8	10
$\|\Gamma\|$ by approximate method	0·000 02	0·000 00	0·000 02	0·000 135
$\|\Gamma\|$ by the computer	0·000 15	0·000 06	0·000 01	0·000 136

It will be seen from the Table that the error is of an oscillatory nature, and becomes smaller as $|\Gamma|$ decreases. It may be concluded that the values of $|\Gamma|$ obtained for the side lobes of the reflection pattern are correct for all practical purposes.

(9.3) Optimum Reflection Pattern

In the design of microwave aerial systems it may be desirable to minimize the side-lobe level for a given width of main beam. Most of the literature on this application is concerned with arrays of discrete elements, but Van de Maas[18] has considered the case of a large number of discrete elements, and the limiting formulae apply to the continuous case. This gives for the best reflection pattern the formula

$$\cos\sqrt{(\omega^2 - a^2)} - \cos\omega$$

where ω is proportional to the authors' l/λ, and a is chosen in accordance with the maximum side-lobe level permitted.

The reflection patterns for the optimum lines obtained by the authors were compared with those obtained from the above cosine formula, and close agreement was found. For example, for the same width of the main lobe the maximum height of the side lobes in the reflection pattern obtained by the cosine formula was found to be 0·000 14 compared with 0·000 15 for the line defined by eqn. (15), showing negligible difference.

The function $P(y)$ associated with the cosine formula involves the Bessel function I, so that the determination of the profile of the tapered centre conductor becomes extremely difficult, if not intractable.

621.372.54 : 621.372.8

Paper No. 2001 R
Mar. 1956

DESIGN OF MICROWAVE FILTERS WITH QUARTER-WAVE COUPLINGS

By G. CRAVEN and L. LEWIN, Associate Member.

(*The paper was first received* 10*th October, and in revised form* 18*th November*, 1955.)

SUMMARY

Resonator diaphragms are constructed from triplets of inductive posts, whose spacings and radii are chosen to prevent the generation of the first five non-propagating modes. This is sufficient to ensure negligible coupling at quarter-wave separation. Measurements on filters at 4 000 Mc/s indicate a satisfactory performance.

INTRODUCTION

Microwave band-pass filters constructed in waveguides are well known and have been fully described.[1] Such a filter consists of any desired number of sections, a section comprising a resonant cavity formed by two obstacles half a guide wavelength apart, followed by a connecting length that is nearly an odd multiple of a quarter guide wavelength long. Thus, quarter-wave connecting lengths would seem to be the logical choice. However, in practice this has the objection that undesirable mutual coupling between the end obstacles of adjacent cavities then exists, and this degrades the performance. For this reason three-quarter-wave couplings are usually preferred, although this leads to filters that are unnecessarily long. In many applications this is a considerable disadvantage and some other solution is desirable.

The purpose of the paper is to show that, by suitably modifying the design of the obstacles employed, the shorter connecting lengths may be retained without introducing noticeable mutual coupling. These modifications in no way affect the usual design procedure, so that where this is required, reference to the original paper[1] should be made. The following work was carried out at 4 000 Mc/s in 2 in $\times$ $\frac{2}{3}$ in waveguide.

(1) OBSTACLE DESIGN

In general, the obstacle used may be of either the capacitive or inductive type; the latter, however, is usually preferable from the manufacturing point of view. It will be seen later that the inductive post, rather than the iris, is the natural choice for the present application.

The mutual coupling that occurs between obstacles when the spacing is small is caused by the higher-order modes they generate. Although evanescent these modes are insufficiently attenuated in the distance allowed and so represent a source of unwanted coupling. An increased spacing is an obvious remedy, but a preferable one is to design the obstacle so that the higher-order modes generated are of small amplitude. This can be achieved if, instead of a single post, several are placed across the guide so that a measure of cancellation occurs.

Fig. 1 illustrates the construction; it shows three posts of equal radius, r, equally spaced across the guide. From the symmetry of construction it is evident that some cancellation of H_{0n} modes ($n > 1$) occurs. Analysis (see Section 7) shows that complete cancellation of the H_{02}–H_{06} modes occurs, i.e. the first higher-order mode generated is the H_{07}, which is very highly attenuated in a short distance.

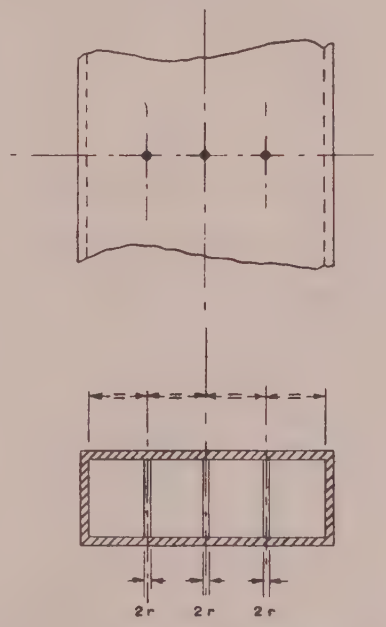


Fig. 1.—Triple-stub obstacle in waveguide.

(2) MEASUREMENT OF OBSTACLE SUSCEPTANCE

The susceptance of the obstacle determines the loaded Q-factor, and so the bandwidth, of the resonant cavity and must be accurately known. The susceptance of the three posts (shown in Fig. 1) as a function of the variables is derived in Section 7, and is

$$B = -\frac{4\lambda_g}{a[\log_e (a/24{\cdot}66r) + 40{\cdot}4a^2/1\,000\lambda^2]} \quad . \quad . \quad (1)$$

where λ is the free-space wavelength, λ_g is the guide wavelength, r is the post radius and a is the guide width.

The values obtained from this formula are compared with experimentally determined values in Fig. 2. Since the method of measurement was somewhat unconventional it will be briefly described after some discussion of accepted methods.

The range of normalized susceptances required in filters lies between 3 and about 15. The most obvious experimental method is to place the guide containing the posts ahead of a matched load and to measure the v.s.w.r. From this the susceptance can easily be calculated. However, the method can involve considerable errors when the v.s.w.r. is from 0·03 to 0·005, the causes having been discussed.[2] Also, guide loss can cause appreciable errors unless corrections are employed.[3] A more suitable method is to place the obstacle between a source of power and a power indicator, such as a thermistor mount, and measure the insertion loss. From this the susceptance can be calculated.

The method employed in the present case was to construct a resonant cavity from two posts, and terminate the waveguide in a good load. The cavity was then tuned to resonance at the test

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Messrs. Craven and Lewin are at Standard Telecommunication Laboratories Ltd.

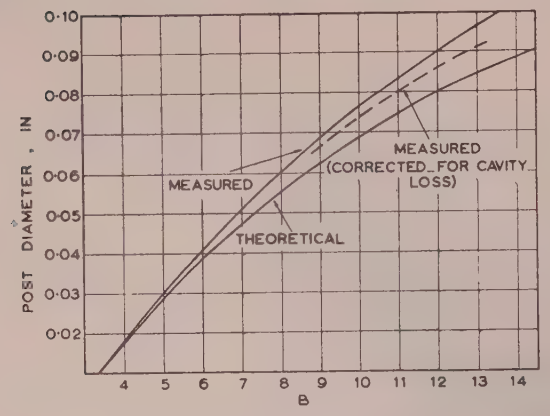


Fig. 2.—Curve of normalized susceptance against post diameter (resonant-cavity method) 4 000 Mc/s.

frequency—a cavity length being chosen so that only a very small tuning-screw insertion was needed—and the bandwidth at the 3 dB points was measured. The Q-factor then being obtained, the normalized susceptance B could be found since the Q-factor of a terminated resonant cavity is related[4] to B by

$$Q \simeq (\tfrac{1}{2}B\lambda_g/\lambda)^2(1 + 2/B^2)\pi[1 - (1/\pi)\arctan(-2/B)] \quad . \quad (2)$$

Although a somewhat indirect method of measuring B, the quantity of ultimate interest, the bandwidth, is directly obtained. Thus correction for the effect of loss on bandwidth in a typical cavity (unplated brass waveguide) is included. However, the main advantage lies in the fact that the inherent accuracy in measuring frequency is high, and that the use of a large amount of attenuation at certain stages of the measurement is avoided.

The value of susceptance obtained by the preceding method includes the effect of loss, which by broadening the bandwidth decreases the apparent value of B. As was discussed above, this is not a disadvantage since bandwidth is the parameter required anyway; but for comparison with the theoretical values obtained from eqn. (1), a true value of B, corrected for cavity loss, is required. The correction may be obtained by measuring the insertion loss of the cavity, from which the equivalent loss resistance in shunt with the load may be found. The modified Q-factor, and thus B, can then be determined. The corrected value of B is shown by the broken curve in Fig. 2; it will be seen that the general agreement between theoretical and experimental values is good.

A 3-section filter embodying the above principles was con-

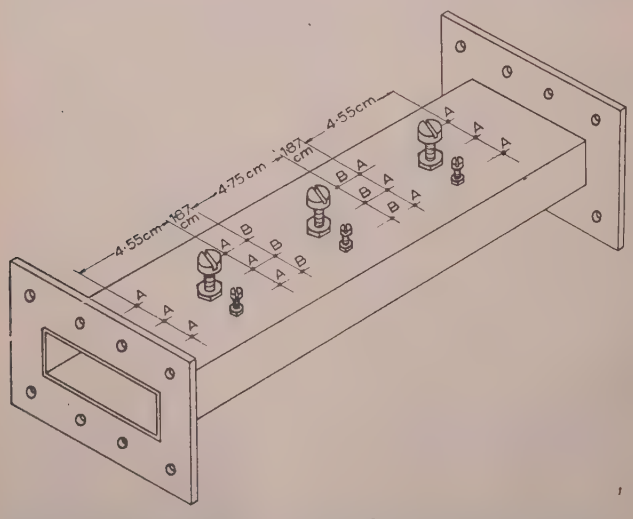


Fig. 3.—Three-section filter in waveguide, using triple-stub obstacle.

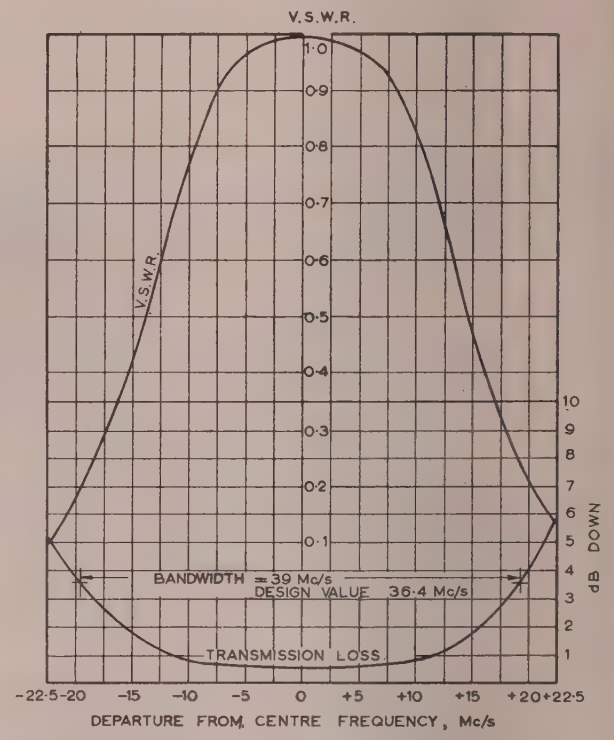


Fig. 4.—Three-section $\lambda_g/4$-coupled filter tuned to design frequency: 4 000 Mc/s.

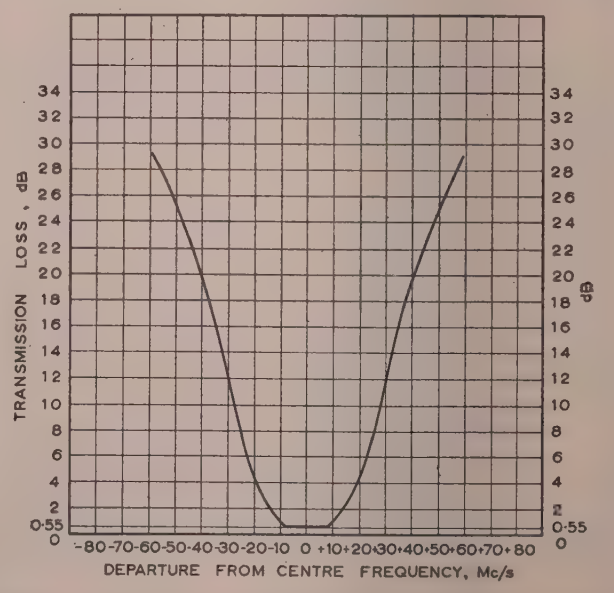


Fig. 5.—Three-section $\lambda_g/4$-coupled filter tuned to design frequency: 4 000 Mc/s.

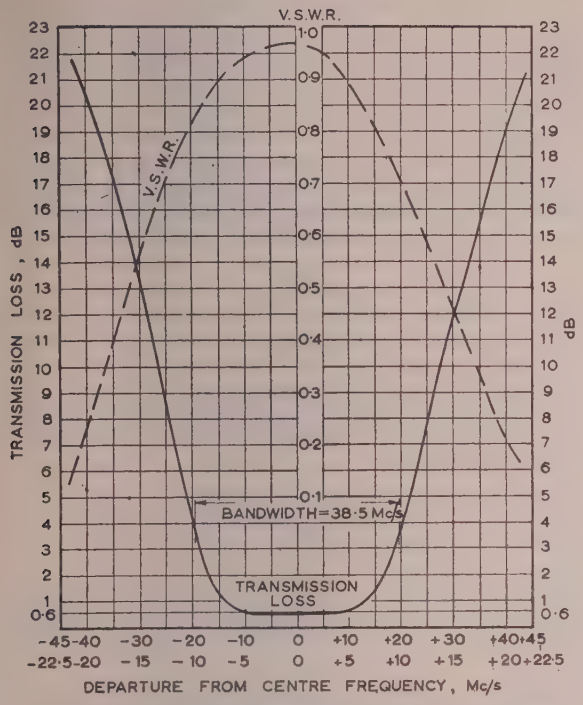


Fig. 6.—Three-section $\lambda_g/4$-coupled filter tuned at 4 050 Mc/s (design frequency + 50 Mc/s).

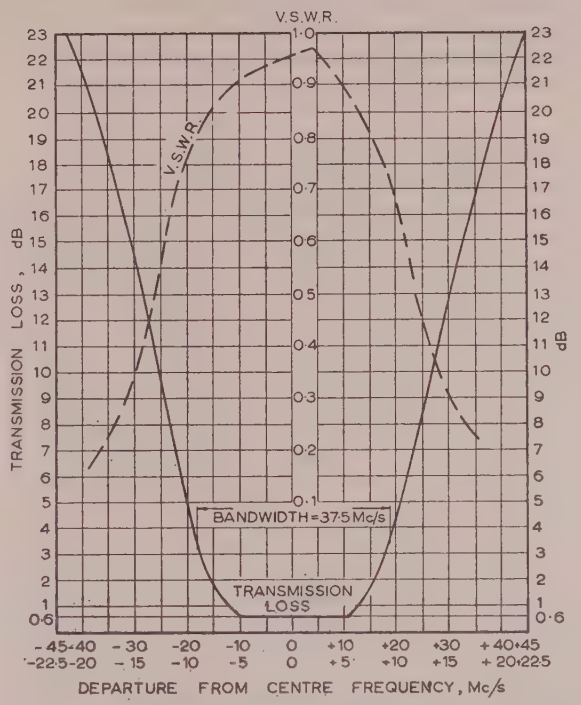


Fig. 8.—Three-section $\lambda_g/4$-coupled filter tuned at 3 950 Mc/s (design frequency − 50 Mc/s).

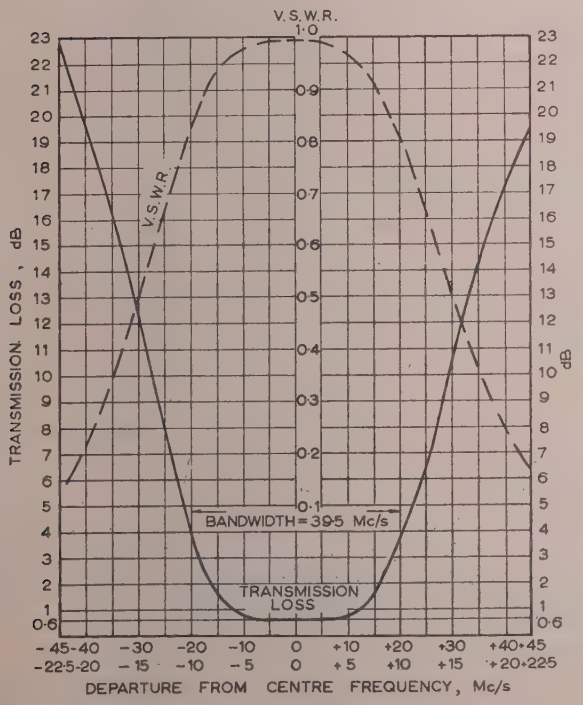


Fig. 7.—Three-section $\lambda_g/4$-coupled filter tuned at 4 100 Mc/s (design frequency + 100 Mc/s).

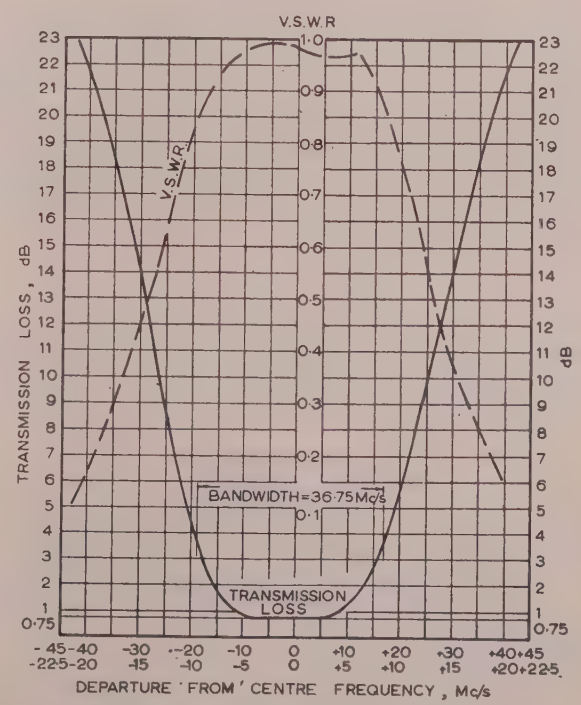


Fig. 9.—Three-section $\lambda_g/4$-coupled filter tuned at 3 900 Mc/s (design frequency − 100 Mc/s).

Figs. 6–9.—Upper scale caption gives transmission loss: lower scale caption gives voltage standing-wave ratio.

structed and tested. The post diameters were determined by the experimental (continuous) curve of Fig. 2. The dimensions and performance at various frequencies are given in Figs. 3–9. In these tests the method of tuning employed was the standard "shift of minimum" technique.[5]

(3) COMPARISON OF FILTERS WITH QUARTER- AND THREE-QUARTER-WAVE COUPLINGS

The saving in length with the shorter couplings can be considerable; a 5-section filter, the greatest number of sections so far tried, designed at 4000 Mc/s (2 in × $\frac{2}{3}$ in guide) and employing quarter-wave couplings, is about 11½ in long, approximately 9 in shorter than its three-quarter-wave counterpart.

The performance of the short filter appears to be at least as good as that of the longer type. The freedom from ripples in the passband v.s.w.r., seen in Fig. 4, is characteristic of the performance at the design frequency. Four- and five-section filters retain this feature. The decreased frequency sensitivity of the shorter couplings is an advantage and should extend the range of tuning frequencies over which acceptable performance can be obtained.

A much smaller diameter is required, for a given susceptance, when three posts replace the usual single post. Electrically, this is an advantage since the ideal of a pure shunt susceptance is more nearly realized; but mechanically it can be a disadvantage since the post diameter can sometimes be quite small. The smallest diameter used so far has been 0·01 in, used in the end cavities (Q-factor ≃ 20) of the 5-section filter. Wire of this size requires some care in assembly; there should, however, be little danger of damage once construction is complete.

(4) CONCLUSION

A number of band-pass filters employing quarter-wave couplings have been built and tested at 4000 Mc/s and good performance has been obtained. Compared with previously constructed filters a considerable reduction in length is possible.

The agreement between theory and experiment is good. Thus, in most cases the equation derived, eqn. (1), may be directly employed for design purposes. However, where the highest standards of performance are required, the post susceptances should be accurately measured.

(5) ACKNOWLEDGMENTS

The authors wish to acknowledge the help of their colleagues in the preparation and measurement of the filters, and their thanks to Standard Telecommunication Laboratories Ltd., for permission and facilities to publish the paper.

(6) REFERENCES

(1) MUMFORD, W. W.: "Maximally Flat Filters in Waveguide," *Bell System Technical Journal*, 1948, **27**, p. 684.
(2) MONTGOMERY, C. G.: "Technique of Microwave Measurements," M.I.T. Radiation Laboratory Series (McGraw-Hill Book Co., 1947), Vol. 11, p. 634.
(3) See Reference 2, p. 507.
(4) See Reference 1, eqn. (32).
(5) RAGAN, G. L.: "Microwave Transmission Circuits," M.I.T. Radiation Laboratory Series (McGraw-Hill Book Co., 1948), Vol. 9, p. 714.
(6) LEWIN, L.: "Advanced Theory of Waveguides" (Iliffe, 1951). Page 25, eqn. (2.8).

(7) APPENDIX

(7.1) Three-Stub Impedance in Rectangular Waveguide

From symmetry considerations, in order to cancel even-order modes, the spacings and radii of the outer stubs must be equal. Let the centre stub be of radius r and the outer stubs of radii R and spaced a distance d from the guide walls.

An incident wave $\varepsilon^{-jk'z}\sin(\pi y/a)$, where $k' = 2\pi/\lambda_g$, excites currents in the stubs. External to the stubs, if their radii are small, the effect is that of filamentary currents I_1 and I_2 at their centres, I_1 in the centre stub and I_2 in each of the outer stubs. Hence the total electric field in the guide can be written[6] as

$$E_x = \varepsilon^{-jk'z}\sin(\pi y/a) + j\sum_1^\infty (k'/\Gamma_m)\varepsilon^{-\Gamma_m|z|}\sin(m\pi y/a)$$
$$\{I_1\sin(m\pi/2) + I_2[\sin(m\pi d/a) + \sin|(m\pi - m\pi d/a)]\}$$

where $\Gamma_m = (m^2\pi^2/a^2 - k^2)^{1/2}$.

In order to make the third and fifth modes zero (the second, fourth and sixth are zero by symmetry) the following two equations must be satisfied:

$$I_1\sin(3\pi/2) + I_2[\sin(3\pi d/a) + \sin(3\pi - 3\pi d/a)] = 0$$
$$I_1\sin(5\pi/2) + I_2[\sin(5\pi d/a) + \sin(5\pi - 5\pi d/a)] = 0$$

These equations give $I_1 = \sqrt{(2)}I_2$ and $d = a/4$.

The electric field has to be zero at the surface of the stubs, i.e. at $z = 0$, $y = (a/2 \pm r)$ and $(a/4 \pm R)$. These conditions give

$$0 = \cos(\pi r/a) + jI_1\sum_1^\infty (k'/\Gamma_m)\sin(m\pi/2 \pm m\pi r/a)$$
$$\left\{\sin(m\pi/2) + \frac{1}{\sqrt{2}}[\sin(m\pi/4) + \sin(3m\pi/4)]\right\}$$

and

$$0 = \sin(\pi/4 \pm \pi R/a) + jI_1\sum_1^\infty (k'/\Gamma_m)\sin(m\pi/4 \pm m\pi R/a)$$
$$\left\{\sin(m\pi/2) + \frac{1}{\sqrt{2}}[\sin(m\pi/4) + \sin(3m\pi/4)]\right\}$$

The expression in braces reduces to

$$[\sin(m\pi/2) + \sqrt{(2)}\sin(m\pi/4)]$$

when m is odd, and is zero when m is even. Adding together the first two equations (containing $\pm r$) and the second two, and restricting m to odd values only, the following two equations are obtained:

$$0 = \cos(\pi r/a) + jI_1\sum_1 (k'/\Gamma_m)\cos(m\pi r/a)$$
$$[\sin^2(m\pi/2) + \sqrt{(2)}\sin(m\pi/2)\sin(m\pi/4)]$$

$$0 = \frac{1}{\sqrt{2}}\cos(\pi R/a) + jI_1\sum_1 (k'/\Gamma_m)\cos(m\pi R/a)$$
$$[\sin(m\pi/4)\sin(m\pi/2) + \sqrt{(2)}\sin^2(m\pi/4)]$$

For odd m the expressions in the square brackets can be seen to differ from each other merely by a factor $\sqrt{2}$, whence it is apparent that these equations are satisfied by taking $R = r$ and

$$I_1 = j\cos(\pi r/a)\Big/\left\{\sum_1 (k'/\Gamma_m)\cos(m\pi r/a)\right.$$
$$\left.[1 + \sqrt{(2)}\sin(m\pi/2)\sin(m\pi/4)]\right\}$$

The reflection coefficient in the dominant mode is given by the $m = 1$ term, and is $\rho = I_1 + I_2[\sin(\pi/4) + \sin(3\pi/4)] = 2I_1$.

A reactance X across the line at the position of the stubs

produces a reflection coefficient given by $X = j(1 + \rho)/2\rho$ and, on substituting for ρ in terms of the expression for I_1,

$$X = \sum_3 (k'/4\Gamma_m) \cos(m\pi r/a)[1 + \sqrt{(2)} \sin(m\pi/2) \sin(m\pi/4)]$$

Now

$$(k'/4\Gamma_m) = (k'/4)(m^2\pi^2/a^2 - k^2)^{-1/2} \simeq (a/2\lambda_g)[1/m + (2a^2/\lambda^2)/m^3]$$

Since r is assumed small the term $\cos(m\pi r/a)$ can be taken as unity except in the summation involving $1/m$, where the cosine term must be retained to prevent divergencies.

Hence $$X = (a/2\lambda_g)\left\{ \sum_3 \left[\frac{\cos(m\pi r/a)}{m} + \frac{2a^2/\lambda^2}{m^3} \right] [1 + \sqrt{(2)} \sin(m\pi/2) \sin(m\pi/4)] \right\}$$

On performing the summation using the result

$$\sum_1^\infty \frac{\cos(2m+1)\theta}{2m+1} = -\cos\theta + \tfrac{1}{2} \log \cot \tfrac{1}{2}\theta$$

and using numerical methods for the $1/m^3$ term, the result quoted in the text for small values of r is found.

DISCUSSION ON "THEORY OF IMPERFECT WAVEGUIDES: THE EFFECT OF WALL IMPEDANCE"*

Dr. V. M. Papadopoulos (*communicated*): An investigation of the effect of finite wall conductance on propagation in cylindrical waveguides has been given previously.† The method employed was essentially the same as that of the present author. The theory was applied to a cylinder of arbitrary cross-section, and general results were given. The existence of stable combination modes was associated with the existence of degenerate modes in a perfect guide, i.e. with the existence of two or more modes which have the same eigenvalue, h_0^2. Some degeneracies are removed by the imperfectly conducting walls, others are not. Thus in a circular waveguide there is degeneracy because two polarizations correspond to a given eigenvalue, and this effect depends on circular symmetry which is not affected by the conductivity of the walls. In the rectangular guide the results apply when all four walls are imperfectly conducting and they show that H_{on}-modes are stable by themselves, while higher-order H-modes and H_{mo}- and all E-modes are stable only in combination.

Dr. A. E. Karbowiak (*in reply*): In his paper Dr. Papadopoulos† has confined his considerations to metallic waveguides (a particular case), and the properties he mentions were discussed in some detail several years ago by Müller.[G]

The purpose of the recent paper[A] was to discuss, with a view to engineering application, basic concepts of surface impedance, to give a few physical examples, and to extend the meaning of the surface impedance so as to enable one to describe any physically realizable homogeneous surface in terms of a single complex quantity, $\boldsymbol{Z}_s$. Thus metal surfaces have been considered to possess a surface impedance[B] $[\boldsymbol{Z}_s = \sqrt{(\omega\mu/\sigma)} \angle 45°]$, which is clearly homogeneous. The term "surface impedance" has also been applied in the past to a dielectric-coated metal surface,[C] and general expressions for its value are available.[D] Even wire grids[E] are known to exhibit a surface impedance characteristic of their configuration, and so do corrugated surfaces.[F] The concept of surface impedance was further developed in the paper to embrace anisotropic surfaces, and there is no reason why it should not be extended to cover even heterogeneous surfaces.

Hitherto, only perfect ($\boldsymbol{Z}_s = 0$) and metallic waveguides $[\boldsymbol{Z}_s = \sqrt{(\omega\mu/\sigma)} \angle 45°]$ have been studied in any detail. In contrast, the analysis of imperfect waveguides as presented in the paper[A] is valid for *any* surface impedance (provided $|\boldsymbol{Z}_s| \ll Z_0$), including anisotropic surfaces, and thus the formulae derived for the coupling, attenuation, and phase-change coefficients of waveguides, as well as the Q-factor and resonant frequency of resonators, are of a wide application. All these results are novel.

In the course of the analysis it was established, among other findings, that all higher order E- and H-modes are unstable in rectangular imperfect waveguides. It is easy to see that it is the reactive part of $\boldsymbol{Z}_s$ that is responsible for the continual oscillation of the energy between the E- and H-modes.

* Karbowiak, A. E.: Paper No. 1841 R, September, 1955 (see **102** B, p. 698).

† Papadopoulos, V. M.: "Propagation of Electromagnetic Waves in Cylindrical Waveguides with Imperfectly Conducting Walls," *Quarterly Journal of Mechanics and Applied Mathematics*, 1954, 7, Part 3.

(A) Karbowiak, A. E.: "Theory of Imperfect Waveguides: the Effect of Wall Impedance," *Proceedings I.E.E.*, 1955, Paper No. 1841 R, September, 1955 (**102** B, p. 698).

(B) Stratton, J. A.: "Electromagnetic Theory (McGraw-Hill Book Co., Inc., 1941), p. 482, *et seq.*

(C) Barlow, H. M., and Cullen, A. L.: "Surface Waves," *Proceedings I.E.E.*, Paper No. 1482 R, November, 1953 (**100**, Part III, p. 329).

(D) Karbowiak, A. E.: "Theory of Composite Guides: Stratified Guides for Surface Waves," *ibid.*, Paper No. 1659 R, July, 1954 (**101**, Part III, p. 238).

(E) Macfarlane, G. G.: "Surface Impedance of Infinite Parallel Wire Grid at Oblique Angles of Incidence," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1523.

(F) Barlow, H. M., and Karbowiak, A. E.: "An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides," *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (**101**, Part III, p. 185).

(G) Müller, R.: "The Stability and Attenuation of Guided Electric and Magnetic Waves of the same Critical Frequency," *Zeitschrift für Naturforschung*, 1949, 4A, June, p. 218.

621.317.755

Paper No. 1931 R
Mar. 1956

AN ULTRA-HIGH-SPEED OSCILLOGRAPH

By F. R. CONNOR, M.Sc., B.Sc.(Eng.).

(*The paper was first received 22nd April and in revised form 2nd August, 1955.*)

SUMMARY

The paper describes the design and construction of an ultra-high-speed oscillograph for observing and studying the build-up of radio-frequency oscillations of a pulsed magnetron, and other phenomena in the millimicrosecond region. It has a linear sweep-speed of 2·5 cm/mμs with a full scan of 5 cm. A special-purpose cathode-ray tube is incorporated having small plate systems. The Y-plate system is designed as a transmission line, which can be used as a resonant or non-resonant line, thus providing a means for narrow-band or broad-band matching. A unit containing short lengths of 70-ohm coaxial cable for delay purposes is included, and there is a calibrating circuit which provides a 1 250 Mc/s timing waveform from a pulsed coaxial-line oscillator which also checks the linearity of the time-base.

(1) INTRODUCTION

One of the many problems to be studied in connection with the stable operation of pulsed magnetrons is the nature of the build-up of oscillations. To assist in this study it was considered that it would be an advantage to see the actual radio-frequency oscillations on a cathode-ray tube, and by their observation to learn more about the behaviour of a pulsed magnetron during the start of the oscillations. This necessitated the construction of an ultra-high-speed oscillograph having a sweep-time in the millimicrosecond region, and the design of a time-base for this oscillograph presented the main problem. Owing to the high speed of sweep required the choice of the type of time-base was extremely small. It was soon realized that the best way of solving the problem was to use some form of constant-current generator to charge or discharge a capacitor.

A fast sweep requires the current to be very large and the capacitor to be as small as possible. Owing to the X-plate sensitivity of the cathode-ray tube used, a scanning voltage of about 600 volts is required. If we assume a total capacitance, C, including the valve, plate and stray capacitances, of 15 $\mu\mu$F, a sweep-time of 2 millimicrosec requires a current of 4·5 amp. The valve required must therefore be capable of supplying a peak current of the order of 5 amp with an output capacity around 10 $\mu\mu$F. Such requirements can be met by either a thyratron or a tetrode. Although thyratrons can be obtained capable of giving peak currents in excess of 5 amp, the jitter in the triggering of the thyratron exceeds the tolerable limit for a scan of 2 millimicrosec. This jitter should be less than 0·02 millimicrosec, which is not beyond the capabilities of a hard valve.

After much consideration of various types of valve it was found that the KT67 or the QQ V06–40 working with high screen and anode potentials satisfied the requirements of high peak current and low output capacitance. Moreover, constant-current operation can be simulated in these tetrodes, for, with decoupling between screen and cathode fairly constant valve currents over a large variation of anode-cathode potential for a given grid-cathode potential can be obtained, and this can be achieved more fully by the use of negative feedback. To measure the peak pulse current attainable with the KT67 and QQ V06–40, measurements were made using the circuits of Figs. 1 and 2.

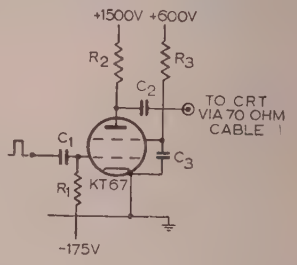


Fig. 1.—Test circuit of tetrode KT67.

$R_1 = 10\,k\Omega$ $C_1 = 0{\cdot}001\,\mu F$
$R_2 = 1$ or $10\,k\Omega$ $C_2 = 0{\cdot}001\,\mu F$
$R_3 = 1\,k\Omega$ $C_3 = 0{\cdot}01\,\mu F$

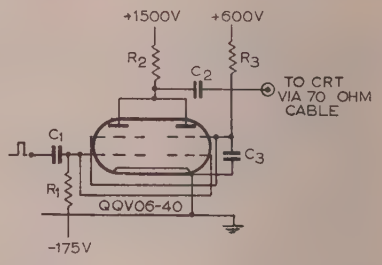


Fig. 2.—Test circuit of double tetrode QQ V06–40.

$R_1 = 10\,k\Omega$ $C_1 = 0{\cdot}001\,\mu F$
$R_2 = 1$ or $10\,k\Omega$ $C_2 = 0{\cdot}001\,\mu F$
$R_3 = 1\,k\Omega$ $C_3 = 0{\cdot}01\,\mu F$

The results obtained for the input and output conditions with a 1-kilohm load are given in Table 1. The input conditions were measured at a high impedance level.

Table 1

PEAK PULSE CURRENTS FROM VALVES KT67 AND QQ V06–40

Input		Output from KT67			Output from QQ V06–40		
Pulse width (half height)	Peak pulse	Pulse width (half height)	Peak pulse	Peak current	Pulse width (half height)	Peak pulse	Peak current
μs	volts	μs	volts	amp	μs	volts	amp
0·1	720	0·2	320	4·6	0·24	400	5·7
0·04	640	0·08	336	4·8	0·20	400	5·7
0·03	560	0·05	384	5·5	0·15	370	5·3
0·01	540	0·04	400	5·7	0·10	370	5·3

(2) CATHODE-RAY TUBE

As it was intended to study the r.f. oscillations of magnetrons oscillating at frequencies up to 3 000 Mc/s, a special-purpose cathode-ray tube had to be designed to enable waveforms of such short wavelength to be displayed. Since transit-time distortion and plate capacitance are the main drawbacks in conventional cathode-ray tubes, the Y-plate system had to be made narrow

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England.

to keep down the transit time to a minimum. Moreover, the plates themselves had to be small to reduce their capacitance. This of course reduces the sensitivity of the plates, which is a drawback when small millimicrosecond pulses are to be examined, and the use of the oscillograph is therefore limited unless some form of distributed amplification is available for obtaining large millimicrosecond pulses. In the case of a pulsed magnetron, however, this is not a serious drawback, since sufficient power is available to give a reasonable deflection.

A further feature of the Y-plate system is its design in the form of a transmission line to reduce transit-time effects and to provide a matched system when it is properly terminated. Alternatively, by extending the line on one side and terminating it with an adjustable short-circuit, use can be made of the standing wave to obtain a voltage maximum at the centre of the Y plates and so provide a larger deflecting voltage.

Of the tube face only a width of 5 cm is usable, since the beam is cut off on either side, and, consequently, the tube might be too small for certain kinds of work. The tube is normally biased to cut off and brightened by a positive pulse on its modulator grid. There are two inputs to the Y plates: one via an *RC* network which applies a balanced shift potential, and one which connects directly via a short piece of coaxial cable. A balanced shift potential is also applied to the X plates so as to bring the most linear part of the sweep across the tube face, and is not normally readjusted. A face plate is fitted to the panel of the tube unit to take a camera.

(3) CALIBRATOR

Owing to the short deflecting time of the time-base, the problem of calibration was difficult. The best way appeared to be to use a sine wave whose frequency could be accurately determined by a wavemeter and to measure time by noting the number of cycles in a given length of sweep. This incidentally provides a linearity check of the time-base as well, through the spacing of the sine wave. Consequently, some form of pulsed oscillator in the 30 cm region was found necessary. However, owing to the low sensitivity of the Y plates it was essential to obtain as much power as possible from the oscillator to get a sufficiently large deflection. It was therefore necessary to use a coaxial-line oscillator capable of giving the required power output.

The oscillator is of the triode type using a coaxial-line circuit with a common grid, and is anode modulated. The modulator (Fig. 7) is a monostable multivibrator with a driver stage and an output stage. It is triggered by a negative pulse of about 30 volts and is locked to the time-base, since the same pulse, suitably delayed, triggers the time-base. The output stage is a cathode-follower employing two valves in parallel to give a large output into a low impedance. The positive pulse of about 500 volts and 0·5 μs duration from the cathode is used to modulate the anode of the calibrating oscillator. To obtain optimum oscillating conditions anode-cathode capacitance can be introduced and adjusted by means of a feedback screw. The grid-cathode and grid-anode lines are tuned to 1 250 Mc/s after having been checked by a wavemeter. The r.f. output is capacitively coupled via a probe and fed by 70-ohm coaxial cable to the Y plates.

(4) DELAY-LINE UNIT

To assist in the visual examination of various parts of a waveform, a delay unit has been provided for delaying either the start of the time-base or that of the signal being examined. Delays of 1·5, 1·0 and 0·5 millimicrosec are available by using short lengths of 70-ohm coaxial cable. An initial delay of approximately 0·1 microsec has to be included owing to the delay in the time-base operation. It has also to be used with the calibrator to bring the oscillations of larger amplitude across the tube face, because the pulsed coaxial-line oscillator takes a finite time to build up to maximum amplitude.

(5) DESCRIPTION OF TIME-BASE CIRCUIT

(5.1) Pulse Generator

The circuit (Fig. 7) is operated by a negative trigger pulse of about 30 volts, which is fed into a buffer cathode-follower, whose grid leak is returned to the cathode to enable it to accept small or large negative triggering pulses. Owing to its high input impedance it presents little loading on the trigger source and its low output impedance is isolated therefrom. The negative pulse from its cathode is differentiated and triggers the monostable multivibrator. For the generation of a time-base whose waveform is not dependent on the shape of the triggering pulse it was decided to use a multivibrator instead of a blocking oscillator; the leading edge of the output of a blocking oscillator is partly dependent on the shape of the driving trigger, although it is more economical in power consumption. The multivibrator used is of the cathode-coupled type, with one valve conducting and the other cut off. The anode and screen of the second valve are taken to higher potentials than is normal so that the valve may deliver a large current when it conducts. The first valve is triggered by the negative pulse from the previous stage, which is differentiated at its grid, and a short pulse output is obtained from the cathode of about 0·25 microsec duration.

The positive pulse from the multivibrator is sharpened and amplified in two driver stages. These are class C amplifiers biased well beyond cut-off with pulse transformers in their anode circuits. The pulse transformers have been designed to give a good h.f. response in order to obtain as sharp a leading edge as possible in the secondary output pulses. The valves are driven by the positive pulses on their grids and produce large negative pulses on their anodes, which are inverted by the secondary windings of the transformers.

(5.2) Time-Base

In order to use the leading edge of the pulse from the previous stage as a linear time-base, it has to be made more linear. This is done by using a cathode-follower constant-current generator, and the principle on which the method is based is that of charging a condenser by a constant current, the voltage across the condenser rising linearly with time.

In this circuit it is the capacitance across the cathode resistor that is being charged. To obtain an ultra-high speed in the rate of charging, this capacitance must be charged by a current of the order of 4·5 amp peak. This has been achieved by using two valves in parallel of suitable type, such as the QQ V06–40. A special low-capacitance heater transformer for this type of valve keeps down to a minimum the capacitance being charged.

It may be remarked here that only one time-base speed is provided. The use of the instrument is thus limited to the kind of specialized work for which it was designed.

(5.3) Push-Pull Output System

Push-pull output to the X plates is obtained from a cathode-follower employing negative feedback whose anode and cathode loads are adjusted to give equal but opposite phase outputs. Distortion of the linear driving waveform is minimized by the large amount of negative feedback. The valve used is capable of supplying the peak current for driving the output capacitances. Since only half the magnitude of the driving waveform is required

at the anode and cathode, this current need be only of the order of 2·5 amp peak, which can be easily supplied by the KT67 valve.

A faster sweep could be obtained with a larger anode resistor and with the cathode earthed, although such a single-ended system leads to some defocusing effects. Alternatively, it was found that a push-pull pulse transformer in the anode gave somewhat inferior results to the cathode-follower circuit, owing to shortcomings in pulse-transformer design.

(5.4) Brightening Pulse

In order to keep down background light, especially for photographic purposes, a narrow pulse not much longer than the time-base duration and with sharp edges is needed. It was found possible to generate a fairly short-duration pulse to satisfy these requirements by differentiating a part of the output from the first driver stage of the time-base unit. The pulse has then to be delayed by a short piece of coaxial cable so as to occur just before the start of the time-base. It is about 30 volts in magnitude and 4 mμs in duration at half height.

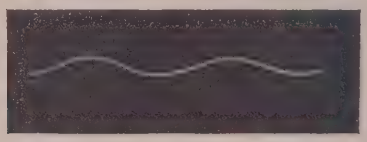


Fig. 3.—Single-stroke 1 250 Mc/s calibrating waveform.

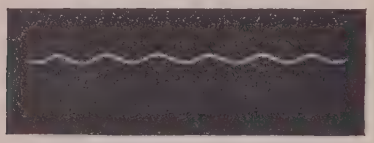


Fig. 4.—Single-stroke r.f. waveform of pulsed 9·1 cm magnetron.

(6) POWER SUPPLIES

To ensure accurate and stable operation of the valves in the time-base-modulator unit and to keep interstage coupling to a

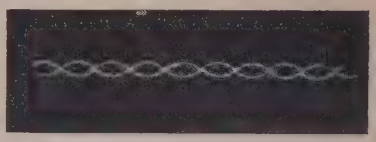


Fig. 5.—Multi-stroke r.f. waveform of pulsed 10·7 cm magnetron. Exposure 0·1 sec.

Fig. 6.—Multi-stroke r.f. waveform of pulsed 9·1 cm magnetron. Exposure 0·1 sec.

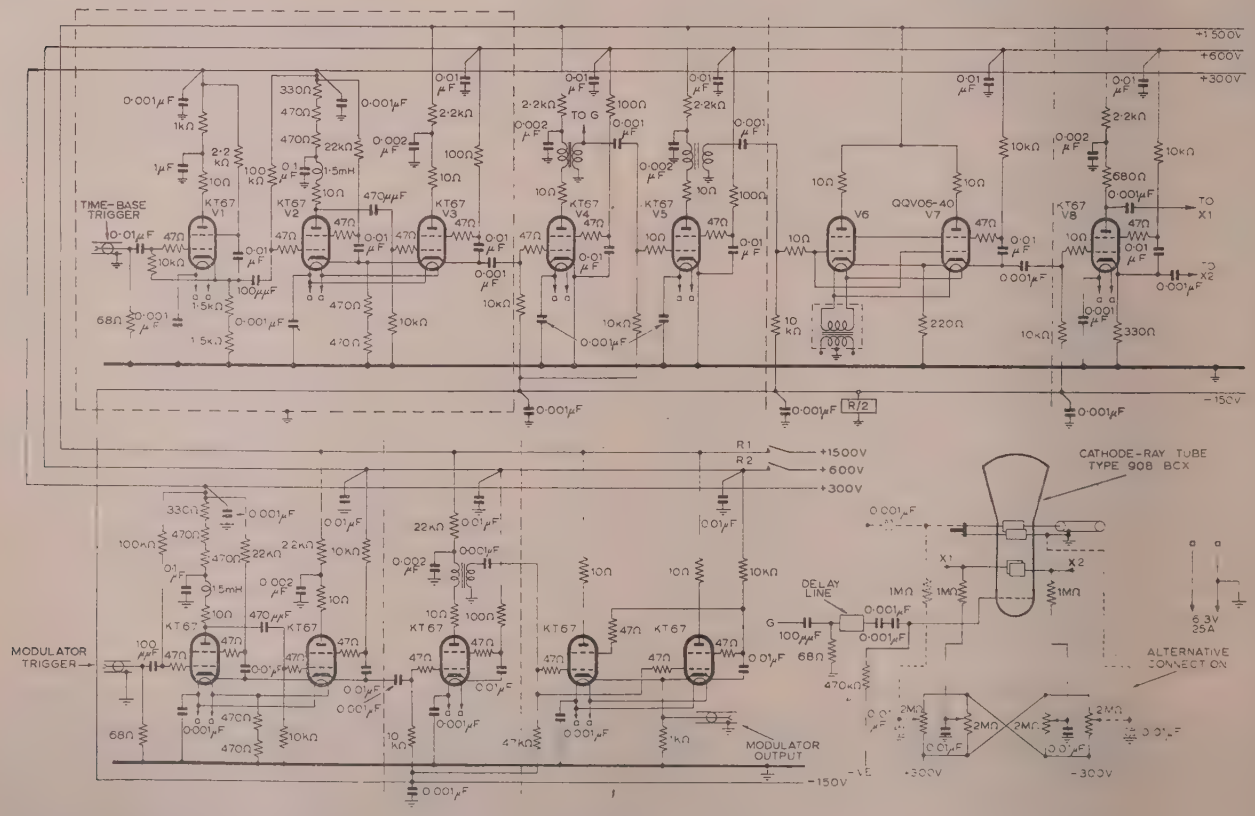


Fig. 7.—Time-base modulator circuit.

minimum, low-impedance stabilized power supplies have been used for the lower voltages. The 50 c/s mains input is filtered through a low-pass filter to prevent r.f. disturbances from other equipment entering the oscillograph via the mains. Additional filtering is provided in the power-supply leads to each stage of the time-base-modulator unit by decoupling with ordinary or feed-through capacitors. The circuits are of conventional design employing a series valve and a control valve. The source of constant-bias reference voltage is provided by a neon stabilizer valve. Output voltages are held constant to within a fraction of one per cent.

(7) PHOTOGRAPHY

The oscillograph is provided with a camera for recording purposes. The camera has an f/1·0 cathode-ray-tube lens with a focal length of 2 in and an object/image ratio of 4 : 1. It uses standard 35 mm film and is mounted direct onto the tube unit on the face plate where it is held in position by four screws. The hood normally in this position is first removed.

The depth of focus of the lens is very small and focusing is effected by means of the lens mount adjustment. It was, however, found impossible to focus the image visually on a ground-glass screen, and this was therefore carried out by taking a series of photographs of the trace, the lens assembly being given one turn after each photograph.

Preliminary sets of photographs were taken to ascertain the optimum conditions for high writing speed.

(8) PERFORMANCE AND USE

The oscillograph has been used in some initial experiments for the study of the build-up characteristics of pulsed magnetrons, and r.f. oscillations from valves operating at 3 000 and 10 000 Mc/s have been displayed by it. The r.f. output is taken from a waveguide via a probe and 70-ohm cable.

Single- and multi-stroke photographs have been taken, and tracings of some typical ones are shown in Figs. 3, 4, 5 and 6.

The tracings in Figs. 5 and 6 appear to suggest that the operation of a pulsed magnetron is not entirely a random phenomenon. Although the sine-wave output varies in phase and an entirely synchronized waveform is not obtained, there is nevertheless an indication of synchronization for one or more phases. This is clearly shown by the fact that the delineation of a phase or two is quite marked against a background of other phases; it may possibly be explained by the suggestion that the pulsed magnetron builds up from time to time in one or more preferred phases, which may occur in some definite order or follow a statistical distribution.

(9) ACKNOWLEDGMENT

The author desires to thank the Board of Admiralty for permission to publish details of the work described in the paper.

(10) REFERENCES

(1) Hughes, V. W., and Walker, R. M.: "Generation of Triangular Waveforms," M.I.T. Radiation Laboratory Series, **19**, p. 259.

(2) Yu, Y. P., Kallman, H. E., and Christaldi, P. S.: "Millimicrosecond Oscillography," *Electronics*, **24**, July, 1951, p. 106.

(3) Lee, R.: "Iron-Core Components in Pulse Amplifiers," *ibid.*, **16**, August, 1943, p. 115.

(4) Rudenberg, H. G.: "Deflection Sensitivity of Parallel-Wire Lines in Cathode-Ray Oscillographs," *Journal of Applied Physics*, **16**, 1945, p. 279.

(5) Bell, J., Gavin, M. R., James, E. G., and Warren, G. W.: "Triodes for Very Short Waves—Oscillators," *Journal I.E.E.*, **93**, Pt. IIIA, 1946, p. 833.

(6) Bereskin, A. B.: "Voltage-Regulated Power Supplies," *Proceedings of the Institute of Radio Engineers*, **31**, 1943, p. 47.

(7) Hill, W. R.: Analysis of Voltage-Regulator Operation," *ibid.*, **33**, 1945, p. 38.

621.373 : 621.3.018

Paper No. 1968 R
Mar. 1956

A COMPARISON OF THE NOISE, AND RANDOM FREQUENCY AND AMPLITUDE FLUCTUATIONS IN DIFFERENT TYPES OF OSCILLATOR

By R. L. BEURLE, B.Sc.(Eng.), Associate Member.

(*The paper was first received 9th July, and in revised form 21st October, 1955.*)

SUMMARY

No oscillator gives a pure sine-wave output. There is always some random noise present in the oscillator circuit, if only that due to thermal agitation. Because of this, noise will appear in the output even if steps are taken to stabilize the output amplitude. Also, the output power is spread over a finite frequency band, and not concentrated at a single frequency.

The nature of the random fluctuation in the output is dependent on the form of feedback employed and the method by which the output amplitude is controlled. The frequency band over which the output power is distributed depends, however, on the relative magnitudes of signal and noise in the circuit rather than on the particular circuit employed. The conclusion is that an oscillator should be operated at as high a level as possible consistent with stability and constancy of the circuit components.

LIST OF SYMBOLS

a, k = Numerical constants.
L, R, C = Inductance, resistance and capacitance.
v, i = Instantaneous values of voltage and current.
$V_{r.m.s.}, I_{r.m.s.}$ = R.M.S. values of voltage and current.
v_L, v_R, v_C = Voltages across inductor, resistor and capacitor.
V_x = Required r.m.s. output voltage measured across the capacitor (or inductor).
$\hat{v}_x$ = Required peak value of output voltage.
i_L, i_R, i_C = Currents in inductor, resistor and capacitor.
i_x = Required output current.
i_d = Total unwanted current.
i_e = Error signal due to lagging servo.
q, q', q'' = Charge on capacitor, and first and second derivatives.
q_0, q_1, q_2 = Charge on capacitor at successive current zeros.
q_∞ = Charge on capacitor at successive current zeros when steady state has been reached.
$q_x = CV_x$.
t = Time.
t_s = Time at which a step-function or delta-function component of noise occurs.
Vdt = Magnitude of the delta function.
F = Bandwidth measured between half-power points, c/s.
δt = Time by which a current zero is displaced.
i_p, v_s = Output of power sources in parallel and series circuits, respectively.
V_s = Supply voltage which is switched from negative to positive, or vice versa, in the circuit considered in Section 3.
$\phi_p(v), \phi_s(i)$ = Output/input characteristic of power sources in parallel and series circuits, respectively.
A = Gain factor with the dimensions of a resistance.
f = Natural frequency of oscillation.
δf = R.M.S. deviation of instantaneous frequency.
$\omega = 2\pi f = \sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)}$
$\omega_0 = 2\pi f_0 = \pm \frac{1}{\sqrt{(LC)}}$
Q = Q-factor $= \frac{\omega_0 L}{R}$
n = An ordinal number.
V_N^2 = (Noise voltage)^2 per unit of bandwidth.
i_N = Total noise current in circuit.
i_n = Noise current due to noise-voltage variations during one half-cycle.
q_n = Charge on capacitor at current zero, due to noise-voltage variation during the previous half-cycle.
q_N = Charge on capacitor at current zero due to noise.
Y = Modulus of admittance of series-tuned circuit.

(1) INTRODUCTION

Oscillators are often classified as feedback oscillators, negative-resistance oscillators, etc., according to the method of introducing power to the circuit. Basically, most of these can be reduced, at least over a limited frequency range, to an equivalent series or parallel tuned circuit containing a source of power. The exceptions are oscillator circuits containing some form of coupled circuit or band-pass filter, but the principles involved are the same even in these circuits and they can be applied to almost any form of oscillator if due allowance is made for the overall response characteristics of the regenerative loop.

The calculations in the rest of the paper apply to the two basic forms of oscillator circuit mentioned. These are shown diagrammatically in Fig. 1. In the parallel tuned circuit, (*a*), P is a current generator the output of which is a function of the voltage across the tuned circuit. The output is fed into the tuned circuit as shown. In the series tuned circuit, (*b*), a source of power generates a voltage which is a function of the current in the circuit. This output voltage is applied in series with the tuned circuit as indicated. In the paper it is shown that the fundamental relationships for the two circuits are analogous, so that only the behaviour of one circuit need be calculated completely.

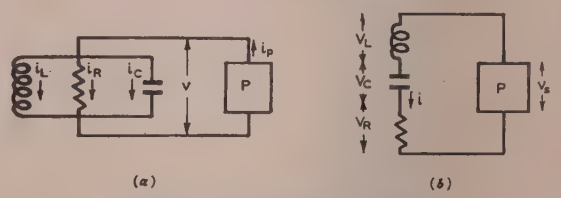


Fig. 1.—Basic oscillator circuits.
(*a*) The parallel circuit.
(*b*) The series circuit.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Beurle is in the Department of Physics, Imperial College, University of London.

If the function relating the output to the input of the power source P is linear and constant, the analysis of behaviour of the circuit is straightforward, and the oscillator produces a narrow band of what is, effectively, noise filtered by a tuned circuit of very high Q-factor. The same applies if the output amplitude is controlled by an a.g.c. system with a time-constant which is long compared with the effective decay time-constant of the tuned circuit including its power source. If, however, the output amplitude is stabilized by limiting the output of P, or is controlled by a rapid a.g.c. servo action on P, it is necessary to analyse the transient behaviour of the circuit during successive half-cycles of the output waveform to determine the frequency and noise content of the output.

(2) FUNDAMENTAL RELATIONSHIPS

The fundamental relationships for both circuits are:

Voltage across capacitor $$v_C = \frac{q}{C} = \frac{1}{C}\int i_C dt + k \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Voltage across resistor $$v_R = Ri_R \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Voltage across inductor $$v_L = L\frac{di_L}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In the parallel circuit, we have

$v_C = v_R = v_L = v =$ input voltage to the power source P

and, in the absence of noise,

$i_C + i_R + i_L = i_p =$ current supplied by power source P

$$= \phi_p(v) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Now, from eqns. (1), (2) and (3),

$$i_C = \frac{dq}{dt} = C\frac{dv}{dt} = Cv'$$

$$i_R = \frac{v}{R}$$

$$i_L = \frac{1}{L}\int v dt + k$$

Substituting these expressions in eqn. (4) gives

$$Cv' + \frac{1}{R}v + \frac{1}{L}\int v dt + k = i_p = \phi_p(v) \quad . \quad . \quad (5)$$

In the series circuit we have

$i_L = i_R = i_C = i =$ current in the series circuit.

Also, in the absence of noise,

$v_L + v_R + v_C = v_s =$ series voltage supplied by power source P

$$= \phi_s(\mathrm{i}) \quad . \quad . \quad . \quad . \quad (6)$$

Now from eqns. (1), (2) and (3),

$$v_L = Li'$$

$$v_R = Ri$$

$$v_C = \frac{1}{C}\int i dt + k$$

Substituting these expressions in eqn. (6) gives

$$Li' + Ri + \frac{1}{C}\int i dt + k = v = \phi_s(i) \quad . \quad . \quad . \quad (7)$$

In the parallel circuit, the source of power provides a current i_p which is a function, $\phi_p(v)$, of the voltage v across the circuit. In the series circuit, the source of power provides a series voltage v_s which is a function, $\phi_s(i)$, of the current i in the circuit. The similarity between eqns. (5) and (7) for parallel and series circuits is evident. If $\phi_p = \phi_s$, the form of the two equations is identical, current in the parallel circuit being the exact analogue of voltage in the series circuit, and vice versa. It is therefore only necessary to make calculations on the behaviour of one circuit, since the results for the one can be readily applied to the other. Hereafter, calculations relating to the series circuit only will be given; analogous calculations may be made for a parallel circuit.

(2.1) Noise

The effect of introducing a small random noise-voltage, V_N, generated in series with the circuit will be considered. The differential equation [eqn. (7)] will thus become

$$Li' + Ri + \frac{1}{C}\int i dt + k = \phi_s(i) + \text{random noise voltage} \quad (8)$$

It will be taken that V_N^2 = voltage^2 per cycle per second of bandwidth, the bandwidth including positive and negative frequencies. In the limit, if only pure thermal noise is present, V_N^2 will have the value $2kTR$, where

k = Boltzmann's constant.

T = Absolute temperature.

R = Resistance across which V_N appears.

(2.2) Methods of Amplitude Control

The purpose of the paper is to study the part played by noise in determining the behaviour of an oscillator, and the dependence of this on the method by which the output amplitude is controlled.

In Section 3, an oscillator without any form of output amplitude control will be considered. The power source will be taken to have a constant linear gain, i.e.

$$V_s = \phi_s(i) = Ai$$

The gain factor A thus has the dimensions of a resistance, and is effectively the negative resistance introduced into the circuit by the power source.

One method of controlling the output of an oscillator is by limiting the output of the power source. If the gain of the power source is high and there is a sharp limiting action, a sine-wave input will give a square-wave output. In Section 4 the gain characteristic $\phi_s(i)$ of the power source will be taken to be of this type. The output is thus a square wave of output voltage of constant magnitude V_s, which is switched alternately from negative ($-V_s$) to positive ($+V_s$) and vice versa at the instant at which the input current passes through zero.

Another method of controlling the output is to apply some form of a.g.c. to the oscillator. The gain factor A may then be a linear function, but will be altered in accordance with variations in the output, so as to bring the output back to the required level. The application of a form of a.g.c. to an oscillator is studied in Section 5. It will be assumed that the gain A remains constant over one half-cycle, from one current zero to the next. The magnitude of A is adjusted by an a.g.c. servo at the beginning of each half-cycle so that, if it were not for noise during the half-cycle, the output amplitude would be restored to a predetermined level by the start of the next half-cycle.

In each Section two assumptions are made in order to simplify the expressions which are derived. The first is that the Q-factor

of the tuned circuit is high, i.e. $R \ll \sqrt{(L/C)}$, so that at the frequency, f_0, given by

$$2\pi f_0 = \omega_0 = 1/\sqrt{(LC)}$$

both $$R \ll \omega_0 L$$

and $$R \ll \frac{1}{\omega_0 C}$$

The second assumption is that the noise current which flows in the simpler LCR tuned circuit owing to the noise voltage V is very much smaller than the oscillatory current due to the output of the power source P.

(3) LINEAR AMPLIFIER OSCILLATOR

(3.1) Calculation of Output

For convenience, the output of the oscillator will be taken to be the voltage across the capacitor. Similar results would be obtained if the output were taken from some other point. From eqn. (8), by putting $i = q'$ we obtain

$$Lq'' + Rq' + \frac{1}{C}q = \phi_s(i) + \text{random noise voltage} \quad . \quad (9)$$

and, since the power source is a linear amplifier,

$$\phi_s(i) = Ai = Aq' \quad . \quad . \quad . \quad . \quad . \quad (10)$$

so that, from eqns. (8) and (9),

$$Lq'' + (R - A)q' + \frac{1}{C}q = \text{random noise voltage} \quad . \quad (11)$$

This is the differential equation representing conditions in a simple series-tuned circuit of inductance L, capacitance C and resistance $(R - A)$, in which the noise voltage is present. For such a circuit the mean-square noise current in the circuit is

$$\overline{i_N^2} = \int_{-\infty}^{\infty} Y^2 V_N^2 df$$

where Y, the modulus of the admittance of the circuit, is given by

$$Y^2 = \frac{1}{\left(\omega L - \frac{1}{\omega C}\right)^2 + (R - A)^2}$$

Integration gives

$$\overline{i_N^2} = \frac{V_N^2}{2L(R - A)}$$

Provided that the Q-factor of the coil is high, this current will be concentrated in a very narrow frequency band about a mean frequency given by

$$\omega_0 = 1/\sqrt{(LC)}$$

The r.m.s. voltage across the inductor L is therefore given by

$$\overline{v_L^2} = \omega_0^2 L^2 \overline{i_N^2} = \pi \frac{\omega_0 L}{(R - A)} V_N^2 f_0$$

The r.m.s. voltage across the capacitor C is given by

$$\overline{v_C^2} = \frac{1}{\omega_0^2 C^2} \overline{i_N^2} = \pi \frac{1}{\omega_0 C (R - A)} V_N^2 f_0 = \overline{v_L^2}$$

It should be noted that v_C, the voltage across the capacitor produced by the amplified random noise-voltage in the circuit, is the output of the oscillator. If the oscillator output is required to have some specified r.m.s. value, V_x, A must be adjusted by some means, once and for all, so that the r.m.s. value of v_C is equal to V_x, i.e.

$$A = R - \pi \omega_0 L \frac{V_N^2 f_0}{V_x^2}$$

The negative resistance, A, is therefore always less than the resistance in the circuit, and the output voltage is in fact merely wide-band noise filtered by a single tuned circuit of high Q-factor (see Fig. 2).

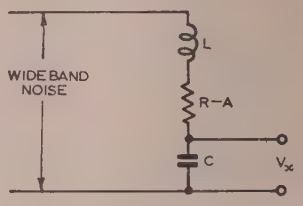


Fig. 2.—Equivalent circuit of linear amplifier oscillator.

(3.2) Frequency Content

The frequency content of the output corresponds to the response of a tuned circuit the Q-factor of which is

$$Q = \frac{\omega_0 L}{(R - A)} = \frac{1}{\pi} \frac{V_x^2}{V_N^2 f_0}$$

Thus the corresponding bandwidth between half-power points is

$$F = \frac{f_0}{Q} = \pi f_0 \frac{V_N^2 f_0}{V_x^2}$$

By using $V_x^2 = \omega_0^2 L^2 \overline{i_N^2}$ this may be rewritten

$$F = \left(\frac{Rf_0}{\omega L}\right) \frac{\left(\frac{1}{2L} V_N^2\right)}{\left(\overline{i_N^2} R\right)}$$

Now it is interesting to note that in the plain LCR circuit without the addition of any power source, a noise current flows which has the mean-square value $V_N^2/2LR$. The noise power dissipated under these conditions is therefore $V_N^2/2L$. When the oscillator circuit is completed by the introduction of the power source the power dissipated becomes $\overline{i_N^2} R$. Thus we express the bandwidth of the output from the complete oscillator circuit in the following form:

$$F = \begin{pmatrix}\text{Bandwidth of plain} \\ LCR \text{ circuit}\end{pmatrix} \times \frac{\begin{pmatrix}\text{Noise power dissipated in} \\ \text{plain } LCR \text{ circuit}\end{pmatrix}}{\begin{pmatrix}\text{Power dissipated in complete} \\ \text{oscillator circuit}\end{pmatrix}}$$

(3.3) Amplitude Fluctuations

The amplitude of the output will be subject to statistical variations following a Rayleigh distribution, even if the r.m.s. output is made equal to V_x. The actual output voltage measured over a short period of time may have any value. This is rather unsatisfactory for most purposes. Generally some form of automatic gain control or other limiting action occurs, even if only accidentally, owing to distortion of valve characteristics or the onset of grid current at high amplitudes. In Section 4 the effect on the output of the power source of introducing a very sharp limiting action is considered.

(4) LIMITING AMPLIFIER OSCILLATOR

(4.1) Characteristics of Power Source

This Section is concerned with an oscillator in which the power source is limited sharply at a voltage $\pm V_s$, so that when

ed from a sinusoidal source the output is a square wave. The magnitude of the output voltage is thus constant ($\pm V_s$) and is switched alternately from negative to positive, and vice versa, in synchronism with the corresponding change in sign of the current in the circuit. The power source thus supplies power to the circuit.

(4.2) Calculation of Output

To calculate the current which flows in the tuned circuit, it is necessary to consider eqn. (8):

$$Lq'' + Rq' + \frac{1}{C}q = \phi_s(i) + \text{random noise voltage}$$

and since $\phi_s(i) = (-1)^n V_s$ where n is the ordinal number of the last current zero,

$$Lq'' + Rq' + \frac{1}{C}q = (-1)^n V_s + \text{random noise voltage} \quad (12)$$

From the general solution, the current during the half-cycle following a reversal of the voltage V_s may be calculated. There are two components to this current. One, corresponding to the complementary function of the general solution, is the oscillatory current in the circuit due to the charge present on the capacitor when the voltage V_s is reversed. This would, in the absence of noise, build up over a number of cycles to a steady state.

The other component, corresponding to the particular integral, is the current due to the noise occurring during the half-cycle. Owing to the presence of the latter component, the length of each half-cycle, as measured between current zeros, will vary. The corresponding frequency variations may be calculated in terms of the ratio of the noise voltage to the oscillatory voltage. Thus the noise voltage produces two forms of disturbance in the output, the current due to the noise and the variation of output frequency which this causes. The variation of the length of the half-cycle also causes a variation in the magnitude of the oscillatory current. However, it may be shown that, unless the Q-factor is low or the noise voltage is high, this variation is very small even compared with the noise current.

(4.3) Oscillatory Current (The Complementary Function)

The circuit will first be considered in the absence of noise, disregarding the particular integral for the moment. If the origin of the time scale is taken to be at a current zero, i.e. if $i = q' = 0$ at $t = 0$, and $n = 0$ at $t = 0$, eqn. (12) will become

$$Lq'' + Rq' + \frac{1}{C}q = + V_s \quad . \quad . \quad . \quad . \quad (13)$$

Then, if the residual charge on the capacitor is q_0, i.e. $q = q_0$ at $t = 0$, it may be shown that the complementary function of the general solution is

$$q = -(CV_s - q_0)\left[\varepsilon^{-(R/2L)t}\cos\omega t + \frac{R}{2\omega L}\varepsilon^{-(R/2L)t}\sin\omega t\right] + CV_s \quad . \quad . \quad . \quad . \quad (14)$$

where

$$\omega = \pm\sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)} \quad . \quad . \quad . \quad . \quad (15)$$

It may be shown by differentiation of eqn. (14) and substitution from eqn. (15) that the current in the circuit is

$$i = \frac{CV_s - q_0}{\omega LC}\varepsilon^{-(R/2L)t}\sin\omega t \quad . \quad . \quad . \quad (16)$$

Starting with zero charge in the condenser at the first current zero, the residual charge builds up exponentially over a large number of cycles to a steady-state residual charge, at subsequent current zeros, of magnitude

$$|q_\infty| = CV_s\frac{1 + \varepsilon^{-(R\pi/2\omega L)}}{1 - \varepsilon^{-(R\pi/2\omega L)}}$$

The sign of the residual charge, of course, changes in alternate cycles from positive to negative and vice versa. The oscillatory current builds up correspondingly to a steady-state value $\pm i$, in positive and negative half-cycles respectively, such that

$$i = \frac{2V_s}{[1 - \varepsilon^{-(R\pi/2\omega L)}]\omega L}\varepsilon^{-(R/2L)t}\sin\omega t \quad . \quad (17)$$

The oscillatory current will not be perfectly sinusoidal because of the exponential factor. The difference appears in the form of odd harmonics, but their total r.m.s. amplitude is only $(14R/\omega L)\%$ of the fundamental. This is small if $R \ll \omega L$. The fundamental component of the current may be found by calculation from the Fourier analysis of a square wave of peak value $\pm V_s$. It is

$$I_{r.m.s.} = \frac{2\sqrt{2}V_s}{R\pi}\frac{\omega}{\sqrt{\left(\omega^2 + \frac{R^2}{16L^2}\right)}}$$

The r.m.s. value of the fundamental component of the voltage across the condenser will be

$$V_x = \frac{2\sqrt{2}}{\pi}\frac{V_s}{\omega CR}\frac{\omega}{\sqrt{\left(\omega^2 + \frac{R^2}{16L^2}\right)}} \quad . \quad . \quad (18)$$

The harmonic content of the voltage across the condenser is less than that of the current because of the decrease in reactance of the condenser with frequency. Noise, which is unavoidably present, produces a further disturbance of this signal, the nature of which will now be investigated.

(4.4) Noise Current (The Particular Integral)

The noise current in the tuned circuit is most easily calculated by the method used in Section 3, giving the r.m.s. noise voltage appearing across the condenser as

$$\frac{I_{N(rms)}}{\omega_0 C} = \frac{\sqrt{\pi}V_N\sqrt{f_0}}{\sqrt{(\omega_0 CR)}} = \sqrt{(Q\pi)}V_N\sqrt{f_0}$$

and the signal/noise ratio

$$\frac{V_x}{\text{r.m.s. noise voltage}} = \frac{1}{\sqrt{(\pi Q)}}\frac{V_x}{V_N\sqrt{f_0}}$$

(4.5) Frequency Variations

In the absence of noise, the oscillatory signal current in the steady state is given by eqn. (17)

$$i = \frac{2V_s}{[1 - \varepsilon^{-(R\pi/2\omega L)}]\omega L}\varepsilon^{-(R/2L)t}\sin\omega t$$

where time t is measured from a current zero.

The time between current zeros is then $t = \pi/\omega$, and this does not vary from one half-cycle to the next. Now, in the presence of a small noise current, each current zero will be displaced slightly to a point at which the oscillatory current and the noise current just cancel. Provided the noise current is very small, the displacement δt will be proportional to the noise current i_n at the current zero, and inversely proportional to the rate of change of the oscillatory current at that point (see Fig. 3).

In order to calculate δt, the current i_n at the end of a half-cycle, due to the noise voltage during the half-cycle, must first be calculated.

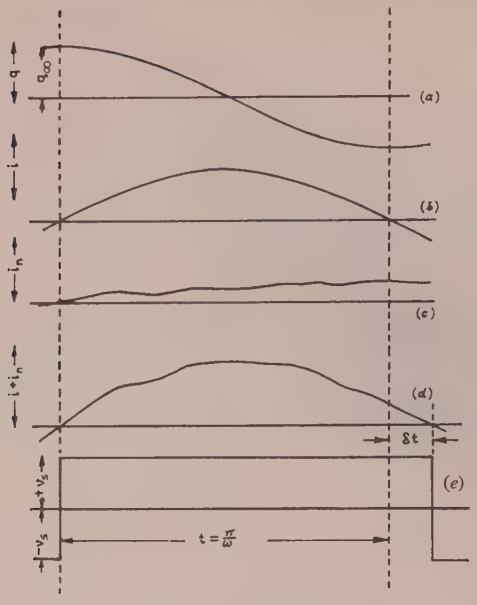


Fig. 3.—Waveforms in limiting amplifier oscillator.

(*a*) Steady-state oscillatory charge (complementary function).
(*b*) Steady-state oscillatory current.
(*c*) Noise current.
(*d*) Sum of waveforms (*b*) and (*c*).
(*e*) Output of power source.

The noise-current waveform will vary from one half-cycle to the next. The particular waveform shown has been exaggerated to illustrate the dependence of δt on the value of i_n at the end of the half-cycle.

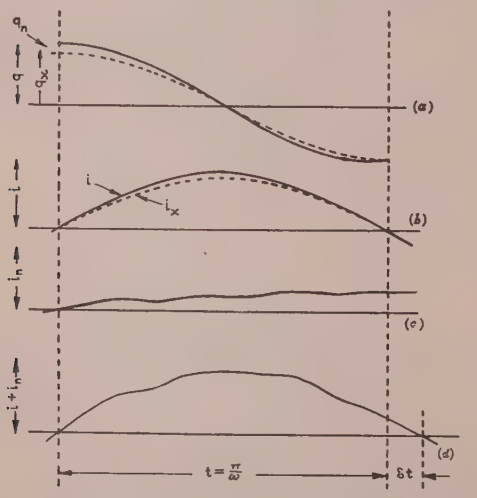


Fig. 4.—Waveforms in oscillator with automatic gain control.

(*a*) Oscillatory charge (complementary function).
---- Desired waveform.
—— Waveform in absence of noise during half-cycle.
(*b*) Oscillatory current.
---- Desired waveform.
—— Actual waveform.
(*c*) Noise current.
(*d*) Sum of waveforms (*b*) and (*c*).

Waveform (*a*) shows how the charge in the capacitor varies throughout the half-cycle in the absence of noise. The charge in the capacitor is assumed, for illustration, to be too high at the start by an amount q_n, and the gain has been correctly adjusted by the a.g.c. servo system to eliminate this discrepancy by the end of the cycle. The presence of noise during the half-cycle will generally cause the charge q to differ again at the end of the half-cycle from the desired value $-q_\infty$.
The value of δt depends on the value of i_n at the end of the half-cycle, as in Fig. 3.

(4.6) Noise Current at End of Half-Cycle

The current i_n can most easily be calculated by considering the random noise voltage (V_N) as a series of delta functions with a Gaussian distribution of amplitude, mean zero, and mean-square magnitude $\overline{(Vdt)^2}$, randomly distributed in time with a mean frequency N, so that $N\overline{(Vdt)^2} = V_N^2$. The current in the circuit due to a step function V at time $t = 0$ has already been calculated (eqn. 16 when $q_0 = 0$ and $V_s = V$); it is

$$i = \frac{V}{\omega L}\varepsilon^{-(R/2L)t}\sin\omega t$$

The current at a time t due to a step function of voltage V at time t_s is therefore

$$i = \frac{V}{\omega L}\varepsilon^{-(R/2L)(t-t_s)}\sin\omega(t-t_s)$$

By differentiation one may obtain the current at a time t due to a delta function (Vdt) at a time t_s; this is

$$\delta i_n = \frac{(Vdt)}{\omega L}\varepsilon^{-(R/2L)(t-t_s)}\left[\omega\cos(t-t_s) - \frac{R}{2L}\sin\omega(t-t_s)\right] \quad (19)$$

By integrating over the half-cycle the contributions of delta functions of mean-square amplitude $\overline{(Vdt)^2}$ randomly distributed in time with a mean frequency N, the r.m.s. current at $t = \pi/\omega$ may be found. If $\overline{N(Vdt)^2}$ is written V_N^2 this is

$$i_{n(r.m.s.)} = \frac{\sqrt{[1-\varepsilon^{-(R\pi/\omega L)}]}}{\sqrt{(2\omega LR)}}V_N\sqrt{\omega} \quad . \quad . \quad . \quad (20)$$

(4.7) R.M.S. Variation in Length of Half-Cycle, and the Resultant Bandwidth

Since it has been assumed that the noise current is very small compared with the current due to the power source, eqn. (17) may here be used for the oscillatory current. The r.m.s. displacement of the zero from its position at $t = \pi/\omega$ may therefore be written

$$\delta t_{r.m.s.} = i_n \Big/ \frac{d}{dt}\left\{\frac{2V_s}{[1-\varepsilon^{-(R\pi/2\omega L)}]\omega L}\varepsilon^{-(R/2L)t}\sin\omega t\right\}_{t=\pi/\omega}$$

Substituting the value of i_n given by eqn. (20) in this, it may be shown that, provided $R \ll \omega L$,

$$\delta t_{r.m.s.} \simeq \frac{\sqrt{\pi}}{4\sqrt{(2)\omega}}\frac{R\pi}{\omega L}\frac{V_N\sqrt{\omega}}{V_s}$$

The displacement has a Gaussian distribution because the current i_n causing it has a Gaussian distribution and the rate of change of current over a small range in the neighbourhood of $t = \pi/\omega$ is constant.

There is thus a Gaussian time shift or phase change introduced each half-cycle. This is effectively frequency-modulation by random noise. For modulation of this type the resultant frequency distribution has been derived by Mr. P. M. Woodward. This derivation is outlined in Section 8. It has the form of a simple resonance in the neighbourhood of the carrier. The bandwidth between 3 dB points on either side is given by

$$F = 4\pi f^3\delta t^2$$

where f is the carrier frequency. Inserting in this equation the value of δt obtained above, we get

$$F = \frac{\pi\omega^2}{64}\left(\frac{R}{\omega L}\right)^2\frac{V_N^2}{V_s^2} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

This may be written

$$F = \frac{1}{2}\left(\frac{R}{\omega L}f\right)\frac{\left(\frac{1}{2}\frac{V_N^2}{L}\right)}{\left(\frac{8}{\pi^2}\frac{V_s^2}{R}\right)}$$

$$= \frac{1}{2}\left(\begin{matrix}\text{Bandwidth of}\\ \text{plain } LCR \text{ circuit}\end{matrix}\right) \times \frac{\left(\begin{matrix}\text{Noise power dissipated}\\ \text{in plain } LCR \text{ circuit}\end{matrix}\right)}{\left(\begin{matrix}\text{Power dissipated at fundamental fre-}\\ \text{quency in complete oscillator circuit}\end{matrix}\right)}$$

This expression is similar to the result quoted in Section 3, but the bandwidth is halved. The bandwidth may be found in terms of the fundamental component of the output voltage by eliminating V_s between eqns. (18) and (21) giving

$$F = \frac{\pi f}{2}\frac{V_N^2 f}{V_x^2} \text{ if } R \ll \omega L$$

This is half the bandwidth of the linear amplifier oscillator; the limiting action, besides reducing fluctuations of amplitude in the output, has reduced the bandwidth to one-half the former value.

(4.8) Effect of Shift of Current Zero on Peak Voltage across Circuit

The effect of the varying length of the half-cycle, on the frequency distribution of the output has been dealt with already. The alteration of the length of the half-cycle will also affect the charge left on the capacitor at the end of the half-cycle. This will produce a noise modulation of the signal amplitude. The resulting random variation in the peak capacitor voltage is added to that due directly to the presence of the noise current. It may be shown that if the Q-factor of the circuit is high, the amplitude of this noise modulation of the signal is $1/2Q$ times the variation of peak condenser voltage due directly to the noise current. It may therefore be neglected in the present problem.

(5) LINEAR AMPLIFIER OSCILLATOR WITH RAPID AUTOMATIC GAIN CONTROL

(5.1) A.G.C. Servo System

This Section is concerned with an oscillator in which the power in the tuned circuit is monitored, and the gain adjusted as necessary in order to restore the power to the desired level. Instantaneous monitoring of the total energy in a tuned circuit would involve measurement and addition of the energy in both condenser and inductance. This would generally be impracticable and the next best thing is to monitor the peak output voltage across the condenser. If this is done, the gain during the half-cycle following a peak may be adjusted so that, if it were not for noise occurring during the half-cycle, the peak output at the end of the half-cycle would be at the correct value. This would constitute the most rapid practical a.g.c. servo system.

(5.2) Calculation of Output

To calculate the output from such a system it is necessary to return to eqn. (11)

$$Lq'' + (R - A)q' + \frac{1}{C}q = \text{random noise}$$

From this the current in the circuit may be calculated. There will be two components to this current. One, corresponding to the complementary function, is the oscillatory current due to the residual charge in the condenser at the start of each half-cycle. The other, corresponding to the particular integral, is the current due to noise occurring during the half-cycle. Owing to the noise current there is a Gaussian variation in the length of successive half-cycles as measured between current zeros. The corresponding frequency spread is calculated as in Section 3.

The noise current also causes a Gaussian variation in the charge which is left in the condenser at the end of each half-cycle. This variation may be calculated from the particular integral. It determines how much the a.g.c. servo system must alter A from its mean value, which will be shown to be R. If A is not equal to R during the following half-cycle, the output departs slightly from the desired sinusoidal form. There are thus two unwanted components in the output: the distortion of the output signal caused indirectly by noise, and the noise itself. The magnitude of the unwanted frequency components corresponding to the distortion of the output may be calculated from the actual output waveform.

(5.3) Oscillatory Current (The Complementary Function)

The circuit will first be considered in the absence of noise, and consideration of the particular integral may be omitted for the present.

By taking a time-scale origin such that $i = q' = 0$ at $t = 0$ and putting $q = q_0$ at $t = 0$, it may be shown that the complementary function is

$$q = q_0\varepsilon^{-[(R-A)/2L]t}\cos\omega t + q_0\frac{(R-A)}{2\omega L}\varepsilon^{-[(R-A)/2L]t}\sin\omega t \quad . \quad . \quad . \quad . \quad (22)$$

where

$$\omega = \pm\sqrt{\left[\frac{1}{LC} - \frac{(R-A)^2}{4L^2}\right]} \quad . \quad . \quad . \quad (23)$$

By differentiation of eqn. (22) and substitution from eqn. (23)

$$i = \frac{dq}{dt} = -\frac{q_0}{\omega LC}\varepsilon^{-[(R-A)/2L]t}\sin\omega t \quad . \quad . \quad (24)$$

In the absence of noise this expression would be a complete description of the series current in the circuit. After a time $t = \pi/\omega$ the current will be zero again and it will be convenient, as in Section 4, to define the beginning and end of each half-cycle as coinciding with current zeros.

(5.4) Effect of Noise (The Particular Integral)

It is simplest to consider first the effect of a single component of the noise voltage present. The current which flows in a circuit of inductance L, capacitance C, and resistance $(R - A)$, due to a delta function of voltage $(V dt)$ in series with the circuit at time t_s is

$$i = \frac{V dt}{\omega L}\varepsilon^{-[(R-A)/2L][t-t_s]} \times \left[\omega\cos\omega(t-t_s) - \frac{R-A}{2L}\sin\omega(t-t_s)\right] \quad (25)$$

This expression is derived in Section 4 [eqn. (19)] from the corresponding differential eqn. (13) for a circuit resistance R. In quoting the result here, R has been replaced by $(R - A)$ to accord with eqn. (11).

(5.5) Bandwidth

The above expression for the current due to a single delta function is identical with the corresponding expression in Section 4, with the exception that R of Section 4 is replaced by $(R - A)$. We may thus make this substitution in the result obtained in Section 4 and, making the same approximations, say that the bandwidth between 3 dB points will be

$$F = \frac{\pi f}{2}\frac{V_N^2 f}{V_x^2}$$

As in Section 4, the bandwidth is half the value derived in Section 3 for an oscillator without any form of amplitude control. The bandwidth may also be expressed, as in Section 4,

$$F = \tfrac{1}{2}\left(\begin{matrix}\text{Bandwidth of}\\ \text{plain } LCR \text{ circuit}\end{matrix}\right) \times \frac{\left(\begin{matrix}\text{Noise power dissipated in plain}\\ LCR \text{ circuit}\end{matrix}\right)}{\left(\begin{matrix}\text{Power dissipated in complete}\\ \text{oscillator circuit}\end{matrix}\right)}$$

This again is half the value derived in Section 3.

(5.6) Amplitude Variation

So far, consideration of the oscillator with automatic gain control has followed the methods of Section 4. Another effect, the disturbance of the signal due to the difference between the actual voltage on the capacitor, at the end of the half-cycle, and the required value $\pm\hat{v}_x$, must now be considered.

The noise current at t, due to a delta function of voltage (Vdt) at t_s, has already been given [eqn. (25)]. By integrating the current due to delta functions of mean square amplitude $\overline{Vdt}^2$ randomly distributed in time with a mean frequency N, the charge on the capacitor, which is the particular integral, may be found. If we integrate this current over the half-cycle we shall get the charge on the capacitor at the end of the half-cycle. If $N\overline{(Vdt)}^2$ is written V_N^2, and $R \ll \omega L$, this charge is found to have the value

$$q_{n(r.m.s.)} = \frac{\pi}{2} C V_N \sqrt{\omega} \quad . \quad . \quad . \quad . \quad (26)$$

If the gain has been correctly adjusted by the a.g.c. servo system in accordance with the value of v_C monitored at the beginning of the half-cycle, the complementary function will have the desired value $\pm q_x = \pm C\hat{v}_x$ at the end of the half-cycle. The particular integral q_n will have a Gaussian distribution about a mean of zero. At the end of each half-cycle the value $\pm(q_x + q_n)$ is thus generally different from the required value, the difference q_n being entirely due to the noise occurring during the half-cycle.

(5.7) Effect of A.G.C. Servo System

At the beginning of each half-cycle the a.g.c. servo system measures the peak voltage across the condenser,

$$\hat{v}_C = \frac{q_0}{C} = \frac{q_x + q_n}{C}$$

The servo system then adjusts A so that the complementary function [eqn. (22)] at the end of the half-cycle ($\omega t = \pi$) is equal to $-q_x$, as shown in Fig. 4. The residual charge, $-q_x$, is the desired value of q_0 for the start of the next half-cycle, and the required value of A is given by

$$-(q_x + q_n)\varepsilon^{-[(R-A)\pi/2\omega L]} = -q_x$$

from which

$$\frac{\pi(R-A)}{2\omega L} = \log_\varepsilon [1 + (q_n/q_x)]$$

and, if

$$\frac{(R-A)}{\omega L} \ll 1$$

then

$$\frac{\pi(R-A)}{2\omega L} \simeq \frac{q_n}{q_x} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Now, since q_n has a Gaussian distribution about a mean of zero, A will have a Gaussian distribution about a mean value of R.

(5.8) Departure from Desired Sinusoidal Current Waveform

The desired series current during each half-cycle is

$$i_x = \frac{q_x}{\omega LC} \sin \omega t$$

where $t = 0$ at the beginning of the half-cycle. Instead of this there will actually be two components of current. These are the current i_n caused by the noise during the half-cycle, and the current given by

$$i = \frac{(q_x + q_n)}{\omega LC} \varepsilon^{-[(R-A)/2L]t} \sin \omega t$$

The difference between i and i_x, the desired series current, is shown in Fig. 4. This difference constitutes an error signal i_e which is due to the lag in the a.g.c. servo system. Its mean-square value may be found by integration over the half-cycle. Provided that $(R - A) \ll \omega L$ the result may be simplified by using the relations in eqns. (26) and (27). It has a value given approximately by

$$\overline{i_e^2} = \frac{\pi}{12} \frac{V_N^2 \omega}{\omega^2 L^2}$$

(5.9) Mean-Square Residual Noise Current

The residual noise current is the current due to the noise occurring during one half-cycle which cannot be eliminated by the a.g.c. servo system described above. It is calculated by integrating over the half-cycle the contributions of delta functions of mean-square amplitude $\overline{(Vdt)^2}$ randomly distributed in time with a mean frequency N. If $\overline{N(Vdt)^2}$ is written V_N^2, and provided that $(R - A) \ll \omega L$, i_n has a value given approximately by

$$\overline{i_n^2} = \frac{\pi}{4} \frac{V_N^2 \omega}{\omega^2 L^2}$$

(5.10) Total Noise

If we use the term noise to cover both the true random noise current present and the error signal resulting from it, we get a total mean-square unwanted noise current given by

$$\overline{i_d^2} = \overline{i_n^2} + \overline{i_e^2}$$

and

$$i_{d(r.m.s.)} = \sqrt{\frac{2}{3}\pi} \frac{V_N \sqrt{f}}{\omega L}$$

The signal/noise ratio will be

$$\sqrt{\left(\frac{\overline{i_x^2}}{\overline{i_d^2}}\right)} \simeq \sqrt{\left(\frac{3}{2}\right)} \frac{V_x}{\pi V_N \sqrt{f}}$$

The combination of noise current and error signal will be essentially random with power present at all frequencies, although there will be a slight predominance in the neighbourhood of the fundamental and of the second harmonic.

(6) CONCLUSIONS

No oscillator generates a pure sine wave. Noise present in the circuit always introduces random phase changes, so that the output power is spread over a narrow frequency band. Other fluctuations will also be present, but the form which these take will depend on the way in which the amplitude of oscillation is controlled. In what has been called the linear oscillator, there is no rapid amplitude control and in consequence the amplitude of the oscillatory output fluctuates with a Rayleigh distribution. This amplitude variation contributes further to the width of the frequency band occupied by the output. The output from this oscillator is, in effect, just noise amplified in a narrow-band amplifier.

The amplitude may be controlled by limiting, or by some form of rapid a.g.c. action. The amplitude of the output signal may then be constant but there will still be noise present in the neighbourhood of the frequency of oscillation. If the amplitude

is controlled by limiting, harmonics will also be present in the output. If the amplitude is controlled by rapid a.g.c. action, noise will also be present at all frequencies but at a lower level.

The numerical results are summarized in Table 1. In the expressions for signal/noise ratio, signal means the steady fundamental component of the output voltage; noise has been taken

Table 1

Type of oscillator	Bandwidth F	Signal/noise ratio	Remarks
Linear . .	$\pi f_0 \frac{V_N^2 f_0}{V_x^2}$	Meaningless	Output is a narrow band of noise.
Limiting	$\frac{\pi}{2} f \frac{V_N^2 f}{V_x^2}$	$\frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{Q}} \frac{V_x}{V_N \sqrt{f_0}}$	Noise is present in region of the fundamental; harmonics are also present.
A.G.C.	$\frac{\pi}{2} f_0 \frac{V_N^2 f}{V_x^2}$	$\sqrt{\left(\frac{3}{2}\right)} \frac{1}{\pi} \frac{V_x}{V_N \sqrt{f}}$	Noise is present at all frequencies

V_x = R.M.S. fundamental output voltage measured across capacitor.
V_N^2 = (Noise voltage)^2 per unit (c/s) of bandwidth.
f = Frequency of output. $f \simeq f_0 = \frac{1}{2\pi\sqrt{(LC)}}$.
$Q = \frac{\omega_0 L}{R}$.

to include any random fluctuation present in the output voltage but does not include steady harmonics. It has also been assumed in the computations for the limiting oscillator and the a.g.c. oscillator that the Q-factor and the ratio of output signal to noise are both high.

The width of the frequency band occupied by the output of the linear oscillator may be regarded as due partly to random phase changes and partly to the sidebands corresponding to amplitude variations. It may be noted that if the variation in output amplitude is removed, by either of the methods considered, the bandwidth is halved. The contributions of phase and amplitude variations to the bandwidth might therefore be regarded as equal, but this idea must not be carried too far since the two are not really separable. Moreover, other methods of controlling the amplitude may not necessarily halve the bandwidth.

The noise fluctuations remaining when the amplitude has been controlled by either method are a small fraction of the signal amplitude. The signal/noise ratio of the oscillator with rapid automatic gain control is greater than that of the limiting oscillator, and its harmonic content is less. The Q-factor appearing in the denominator of the expression for the signal/noise ratio of the limiting oscillator may be slightly misleading, but it must be remembered that, for a given power source, V_x will be proportional to Q. Also, V_N^2, if it is due to thermal noise, will be proportional to the resistance in the circuit. Thus, as one might expect, a high Q-factor will improve the signal/noise ratio.

Both the bandwidth and the noise power relative to the signal power are inversely proportional to $V_x^2/V_N^2 f_0$. This ratio should therefore be as high as possible consistent with stability and constancy of the components.

The control of amplitude in actual oscillators may not correspond exactly to either of the idealized conceptions which have been considered. In spite of this, such oscillators appear to behave according to the expressions in Table 1. Some form of automatic gain control or limiting action is usually present whether or not it is deliberately introduced, although generally the control of amplitude is not as rapid as in the idealized forms. In these circumstances the noise modulation of the output will be greater. Ultimately, if the time-constant of the amplitude control is long enough to be comparable with the decay time-constant of the complete circuit, the amplitude control ceases to be effective. The oscillator then acts as a linear oscillator producing a narrow band of random noise. There is little danger of this happening with the normal type of valve oscillator in which the inherent decay time-constant of the circuit is long compared with the time-constant of the amplitude control.

(7) ACKNOWLEDGMENTS

The author would like to thank his former colleagues at the Radar Research Establishment, Mr. S. W. Noble and Mr. P. M. Woodward, the former for bringing the problem to his attention, and the latter for reading the paper in draft form and for supplying the information outlined in the Appendix. Acknowledgment is made to the Controller of H.M. Stationery Office for permission to publish the paper.

(8) APPENDIX

The broadening of the spectral line resulting from the introduction of small random phase changes each half-cycle comes about as follows.

At every half-cycle there is a small random increment of phase, having an r.m.s. value of $2\pi f \delta t$. After n half-cycles, the total phase error will build up by a factor $\sqrt{n}$, in the usual manner of random increments with zero mean. When the accumulated error is of the order of π, it can be said that the waveform has lost its coherence. The time taken for this to happen is n half-cycles, and the relation may be written

$$\sqrt{(n)} 2\pi f \delta t = \pi \text{ approximately,}$$

so that

$$n = 1/(2f\delta t)^2$$

The time interval for which approximate coherence of phase is maintained is thus given by

$$T = \frac{1}{2f} \frac{1}{(2f\delta t)^2}$$

and the broadening of the spectral line resulting from this finite coherence time is given approximately by $1/T$.

However, it cannot be expected that such a rough argument will give the correct numerical constant. Recourse must be made to a detailed analysis of the auto-correlation function of the waveform. Such analysis shows that, provided δt is sufficiently small, the spectrum has the form of a simple resonance, and the bandwidth between 3 dB points is given by

$$F = 4\pi f^3 \delta t^2$$

It will be seen that this formula agrees with the one derived more simply, apart from the numerical constant.

621.376.3 : 621.3.018.78

Paper No. 2006 R.
Mar. 1956

AN EXTENDED ANALYSIS OF ECHO DISTORTION IN THE F.M. TRANSMISSION OF FREQUENCY-DIVISION MULTIPLEX

By R. G. MEDHURST, B.Sc., and G. F. SMALL, B.Sc.(Eng.), Graduate.

(*The paper was first received 29th July, and in revised form 23rd November*, 1955.)

SUMMARY

When a frequency-division-multiplex telephony signal is subjected to some non-linear process, intermodulation distortion is generated, being usually audible as an unintelligible noise. In f.m. systems an important source of such non-linear distortion is to be found in small echoes of the main r.f. signal. Albersheim and Schafer have investigated theoretically and experimentally the distortion due to a single echo, but the analysis applied only to limited ranges of the available parameters, and the agreement with experiment was not always good. It now appears that, by a combination of analytical and semi-empirical methods, Albersheim and Schafer's results can be extended to enable prediction of distortion levels for most cases of practical interest (in the absence of pre-emphasis). Most of the discrepancy between theory and the measurements of Albersheim and Schafer can also be accounted for, the formulae being in good agreement with measurements carried out by the present authors.

LIST OF SYMBOLS

[To facilitate comparison with the formulae of Albersheim and Schafer, the symbols used in Reference 1 are given in parentheses.]

	ω_c	= Carrier frequency, rad/sec.
ω_M	(M)	= Modulation frequency, rad/sec.
ω_m	(p)	= Frequency of one component tone of ω_M, rad/sec.
$\hat{\omega}_m$	(p_m)	= Maximum value of ω_m, rad/sec.
$\check{\omega}_m$	(p_0)	= Minimum value of ω_m, rad/sec.
ω_D	(s)	= "Peak" frequency deviation (taken as 11 dB above the r.m.s. deviation), rad/sec.
ϕ_{ω_m}	(ϕ_p)	= Random phase angle associated with the tone of frequency ω_m, rad.
	r	= Ratio of echo amplitude to amplitude of main signal.
	τ	= Echo delay time, sec.
V_D	(D)	= Amplitude of a distortion component of frequency ω_m, volts.
V_{Ds}	(D_s)	= Amplitude of a distortion component of frequency ω_m, when the echo is phased for maximum distortion, volts.
V_{Dc}	(D_c)	= Amplitude of a distortion component of frequency ω_m, when the echo is phased for minimum distortion, volts.
V_{D2}	(D_2)	= Amplitude of a second-order distortion component of frequency ω_m, volts.
V_{D3}	(D_3)	= Amplitude of a third-order distortion component of frequency ω_m, volts.
V_U	(S)	= Amplitude of the undistorted base-band component of frequency ω_m, volts.
ϕ_D	(μ_d)	= Phase distortion due to echo, rad.
ϕ_{Mt}	(μ_t)	= Phase modulation $= \int \omega_M dt$, rad.
$\phi_{M(t-\tau)}$	$[\mu_{(t-\tau)}]$	= Delayed phase modulation, rad.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

(1) INTRODUCTION

It is common practice in line telephony to transmit many conversations simultaneously along the same coaxial cable by the use of frequency-division multiplex (f.d.m.). In this system a frequency band, which usually extends from a small fraction, to a number, of megacycles per second, is divided into consecutive 4 kc/s sections, each of which is occupied by a speech channel, suitably translated in frequency. In situations where a cable link may present great practical difficulties and be excessively costly, increasing use is being made of microwave radio systems, in which the same multiplex signal modulates a high-frequency carrier. An important design consideration governing such radiotelephony systems concerns the minimization of intermodulation distortion (usually appearing as an unintelligible noise) due to the passage of the modulated carrier through non-linear circuits. In the paper, frequency-modulation systems are considered; here the problem of the evaluation of distortion due to non-linearity is of considerable difficulty, for it is a peculiarity of frequency-modulation problems that they increase in complexity very rapidly as the number of tones in the modulating signal increases above a very small number, and even for the simplest case of a single tone, much work[2–8] has led to solutions that still retain substantial complication.

It is evident that if the present problem is to be made tractable the multiplex signal must be represented in an idealized form. It has been shown[9] that for a sufficiently large number of channels the statistical properties of the signal approach those of random noise. Thus, assuming that all the channels, when working, are at approximately the same level, it is natural to replace the signal, for calculation and test purposes, by a band of white noise at an r.m.s. level equal to that not exceeded by the system over some large proportion of the time (say 99%).

With such a modulating signal, formulae have been derived[1] for certain limiting values of the available parameters, in the important case of distortion due to an echo (generated, for example, by a mismatched cable or by multi-path transmission). It has been pointed out[10] that these formulae do not cover certain ranges of conditions of considerable practical interest, and, as will be shown below, they are not in uniformly good agreement with experiment. It is the purpose of the paper to show how the results given in Reference 1 may be corrected and extended (by a combination of analytical and semi-empirical methods) to cover most cases of echo distortion likely to arise.

(2) MEASUREMENT TECHNIQUE

Since one objective of the paper is to attempt to resolve discrepancies between Albersheim and Schafer's theory[1] and experiment, substantial use is made of their measurements where they cover an adequate range of conditions. However, further measurements were found necessary to supplement this earlier work, and a brief account of the experimental procedure will be given.

Distortion measurements were made on a system consisting of a transmitter and a receiver connected so as to permit the

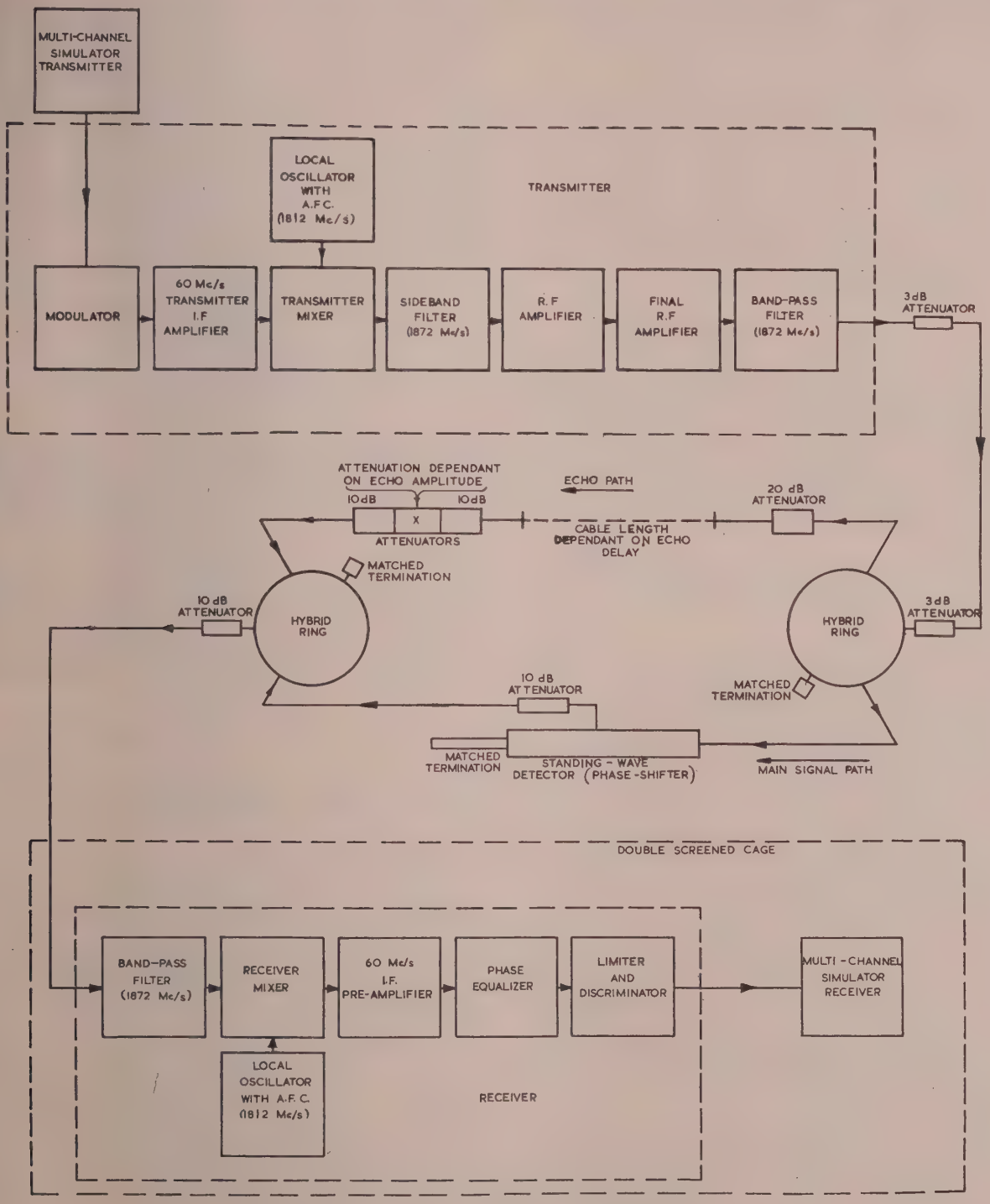


Fig. 1.—Block diagram of distortion-measuring equipment.

introduction of an echo of known characteristics. A block diagram of the equipment is shown as Fig. 1; the transmitter and receiver were units from a 240-channel 2000 Mc/s trunk radio system.

The measurements were made with multi-channel simulator equipment, which generates a band of noise extending between the highest and lowest multi-channel frequencies and is flat except for a narrow, substantially noise-free, gap located at some desired position.

The present work was carried out at radio rather than inter-

mediate frequency (as was done in Reference 1), since in a trunk radio system it is in the r.f. path that echo distortion is expected to be severe.

(2.1) Transmitter and Receiver

The transmitter delivered an output power of about 1 watt at a frequency of 1 872 Mc/s. The modulating noise simulating the 240-channel f.d.m. signal covered the frequency range 60–1 052 kc/s.

The receiver had an equivalent noise power, referred to the input, of about −115 dBW. After demodulation the signal was fed into the receiver unit of the multi-channel simulator. The measurements were made in a channel (centring on 1 040 kc/s) close to the upper end of the base band.

(2.2) Method of Introducing the Echo

An echo was required having an amplitude which was a known fraction of the main signal amplitude, and with a known delay; the phase angle between the echo and the main signal had to be controllable. These conditions were fulfilled by dividing the transmission line connecting the transmitter and the receiver into two paths, which were later recombined. One of these paths carried the main signal and incorporated a phase-shifter; the other carried the echo, which was delayed by the insertion of a length of cable and then attenuated to obtain the required echo amplitude.

To reduce unwanted echoes to negligible proportions, all cable lengths other than that of the delay cable were kept as short as possible, and reflections were minimized by the use of well-matched attenuators.

Coaxial-line hybrid rings were used to divide and combine the two paths, these devices having much better impedance characteristics than the simpler T-junction.

The phase-shifter in the main signal path was a standing-wave line, the attenuation of this unit (about 30 dB) being allowed for by additional attenuation in the echo path. A continuous and known phase variation was obtainable by movement of the probe carriage. The termination of the standing-wave line in such an arrangement must be well matched to the line, otherwise appreciable standing waves result, causing a change in the attenuation of the main signal path with change of phase and hence a change in the echo/main-signal ratio.

The cable used to provide the echo delay had an attenuation of about 0·024 dB/ft at the test frequency of 1 872 Mc/s, and a velocity ratio of 0·96. The total attenuation between the transmitter and the receiver was about 60 dB. To prevent unwanted signal pick-up the flexible coaxial cables used were provided with an additional insulated outer conductor. The attenuators used to ensure isolation and good impedance matches between the various units consisted of short lengths of coaxial line containing a lossy dielectric of resin loaded with iron dust, the ends of the dielectric being tapered to obtain a good impedance match (v.s.w.r. $\gtrsim 1{\cdot}15:1$).

Desired echo amplitudes were obtained as follows. The echo and the main signal were first made equal by removing the attenuator marked X in Fig. 1 and adjusting the standing-wave-line phase-shifter for zero power in the receiver, the attenuation of the standing-wave line being varied if necessary by altering the depth of penetration of the probe. Then, insertion of an attenuator of known value in the echo path at X gave the required echo/signal ratio.

(2.3) Background Noise Level

Sources of distortion noise in the equipment described above are: (*a*) the measuring equipment itself; (*b*) thermal noise in the mixer and first stages of the system receiver; (*c*) non-linearity of the system units (i.e. modulator, amplifiers and demodulator); (*d*) the introduced echo.

Contribution (*a*) was negligible, and (*b*) was also made negligible by using a high level of input signal into the system receiver. This corresponded to an i.f. noise/signal ratio of about −55 dB. Contribution (*c*), when measured without an echo, resulted in a distortion/signal ratio of −50 dB.

Measured distortion/signal ratios due to an echo were corrected for residual system noise on the assumption that distortions from the two sources added on a power basis. Because of the residual noise it was not possible to measure echo distortion/signal ratios less than about −55 dB.

(3) SUMMARY OF RESULTS DUE TO ALBERSHEIM AND SCHAFER

The frequency modulation is written in the form

$$\omega_M = a \sum_{\omega_m=1}^{\hat{\omega}_m} \cos(\omega_m t + \phi_{\omega_m}) \quad . \quad . \quad . \quad (1)$$

where ω_m increases in steps of 1 radian/sec. In this representation the minimum modulating frequency is taken as being negligibly small compared with $\hat{\omega}_m$. If the peak frequency deviation is considered as being 11 dB above the r.m.s. value, it is easy to show that

$$a = 0{\cdot}4\omega_D\hat{\omega}_m^{-0{\cdot}5}$$

Albersheim and Schafer's results[1] for distortion due to a single echo, such as that produced by a mismatched aerial, can be summarized as follows:

The ratio of the distortion component having frequency ω_m to the undistorted base-band component at the same frequency can be written in the form

$$\frac{V_{Ds}}{V_U}\sin\omega_c\tau + \frac{V_{Dc}}{V_U}\cos\omega_c\tau \quad . \quad . \quad . \quad . \quad (2)$$

V_{Ds}/V_U and V_{Dc}/V_U are voltage ratios which are mutually incoherent, i.e. their magnitudes are to be added on a power basis.

When τ is sufficiently small

$$V_{Ds}/V_U \to V_{D2}/V_U = 0{\cdot}20r\tau^2\omega_D\omega_m\sqrt{[1-\tfrac{1}{2}(\omega_m/\hat{\omega}_m)]} \quad (3)$$

and $$V_{Dc}/V_U \to V_{D3}/V_U = 0{\cdot}028r\tau^3\omega_D^2\omega_m\sqrt{[1-\tfrac{1}{3}(\omega_m/\hat{\omega}_m)^2]} \quad (4)$$

When τ is sufficiently large, V_{Ds}/V_U and V_{Dc}/V_U approach limiting values which are numerically equal, but incoherent, so that the distortion no longer depends on the echo phase. In this case, when $\omega_D/\hat{\omega}_m \gg 1$,

$$\frac{V_{Ds}}{V_U} = \frac{V_{Dc}}{V_U} \to \frac{V_D}{V_U} = 5{\cdot}8r\left(\frac{\hat{\omega}_m}{\omega_D}\right)^{0{\cdot}5}\left(\frac{\omega_m}{\omega_D}\right)\exp -2{\cdot}88\left(\frac{\omega_m}{\omega_D}\right)^2 \quad (5)$$

τ is regarded as being small [as required for eqns. (3) and (4)] when $\tau\omega_D \ll 1$ and $\tau\hat{\omega}_m \ll 1$, and as being very large [as required for eqn. (5)] when the reverse relationships hold.

In the top channel, where worst distortion is expected, eqns. (3), (4) and (5) become

τ small

$$\frac{V_{Ds}}{V_U} = \frac{V_{D2}}{V_U} = 0{\cdot}14r\tau^2\omega_D\hat{\omega}_m \quad . \quad . \quad . \quad . \quad (6)$$

$$\frac{V_{Dc}}{V_U} = \frac{V_{D3}}{V_U} = 0{\cdot}023r\tau^3\omega_D^2\hat{\omega}_m \quad . \quad . \quad . \quad . \quad (7)$$

τ large

$$\frac{V_D}{V_U} = 5{\cdot}8r\left(\frac{\hat{\omega}_m}{\omega_D}\right)^{1{\cdot}5}\exp -2{\cdot}88\left(\frac{\hat{\omega}_m}{\omega_D}\right)^2 \quad . \quad . \quad (8)$$

In the present paper attention is confined to the top channel, so that primary concern is with eqns. (6)–(8). Eqns. (6) and (8) have been quite extensively verified[1] over the ranges in which they are valid. Eqn. (7) would also appear to be verified by a set of measurements plotted in Fig. 4 of Reference 1. However, eqn. (4), on which eqn. (7) is based, was incorrectly derived in Appendix 1 of this reference, a factor 3/2 having been omitted, and in addition there was an error in plotting the formula as given. The final result is that, over the region in which eqn. (7) was expected to hold, the measured distortion levels are about 9 dB lower than those predicted (see Fig. 4).

It appears, finally, that in order to complete the work of Reference 1, it is necessary to develop formulae expressing the distortion when τ is neither large nor small, to remove from eqn. (8) the restriction on $\omega_D/\hat{\omega}_m$ and to clarify the discrepancy between eqn. (7) and the measurements in Reference 1.

(4) RELATION BETWEEN DISTORTION AND ECHO AMPLITUDE FOR SMALL DELAY TIMES

Eqns. (6)–(8) require that the echo amplitude, r, should be small. In the present Section this restriction will be removed in the short-delay case. It will be shown that eqn. (7), as it stands, was not in good agreement with measurement because the echo strength used (−15 dB) was not sufficiently small for the simple formula to apply.

Eqns. (3) and (4) were obtained in Reference 1 by approximating to first order in r at an early stage of the analysis. The general expression for the distortion/signal ratio, valid for any r, is obtained in Section 10.1, in the case of small τ. It is

$$0{\cdot}20\tau^2\omega_D\omega_m\left[\frac{r(1-r^2)\sin\omega_c\tau}{(1+2r\cos\omega_c\tau+r^2)^2}\right]\sqrt{\left[1-\frac{1}{2}\left(\frac{\omega_m}{\hat{\omega}_m}\right)\right]}$$

$$+\,0{\cdot}028\tau^3\omega_D^2\omega_m\sqrt{\left[1-\frac{1}{3}\left(\frac{\omega_m}{\hat{\omega}_m}\right)^2\right]}$$

$$\left\{\frac{r(1-r^2)[\cos\omega_c\tau+4r-2r(\cos\omega_c\tau)^2+r^2\cos\omega_c\tau]}{(1+2r\cos\omega_c\tau+r^2)^3}\right\} \quad . \quad . \quad . \quad . \quad (9)$$

where $\omega_c\tau$ is the phase angle of the echo relative to the main signal. As r decreases, it is seen that eqn. (9) degenerates to eqn. (2), V_{Ds}/V_U and V_{Dc}/V_U having the values given in eqns. (3) and (4).

In the degenerate case described by eqns. (2)–(4) the distortion is minimum when the echo is in phase or in antiphase with the main signal, and maximum when it is in quadrature. When r cannot be considered very small, minimum distortion still occurs when the echo is in phase, but not when it is in antiphase, and for maximum distortion the phase angle is displaced from quadrature towards the antiphase position. To a good approximation, minimum distortion is given by

$$\frac{V_{D3}}{V_U}=0{\cdot}028\frac{r(1-r)}{(1+r)^3}\tau^3\omega_D^2\omega_m\sqrt{\left[1-\frac{1}{3}\left(\frac{\omega_m}{\hat{\omega}_m}\right)^2\right]}$$

$$=0{\cdot}023\frac{r(1-r)}{(1+r)^3}\tau^3\omega_D^2\hat{\omega}_m \quad . \quad (10)$$

in the top channel. The expression for maximum distortion is a little more complicated. It is

$$V_{D2}/V_U=0{\cdot}20\mathrm{F}(r)\tau^2\omega_D\omega_m\sqrt{[1-\tfrac{1}{2}(\omega_m/\hat{\omega}_m)]}$$

$$=0{\cdot}14\mathrm{F}(r)\tau^2\omega_D\hat{\omega}_m \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

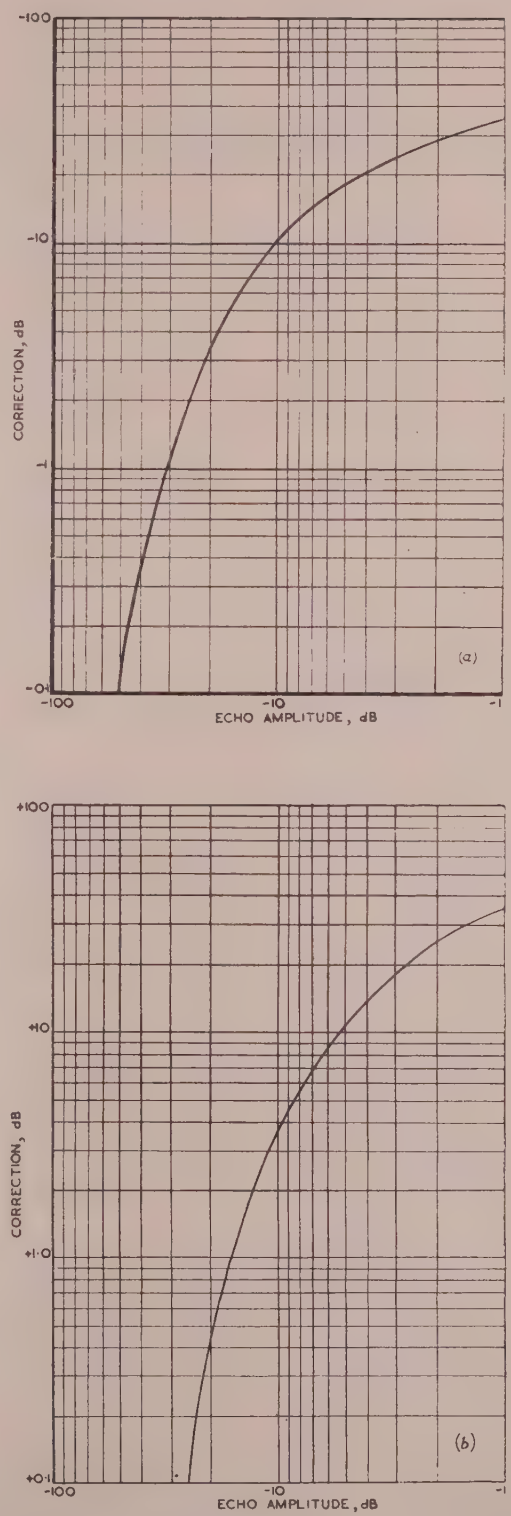


Fig. 2.—Correction to small-echo formula for distortion/signal ratio in top channel.

Delay time small.
(*a*) Echo phased for minimum distortion.
(*b*) Echo phased for maximum distortion.

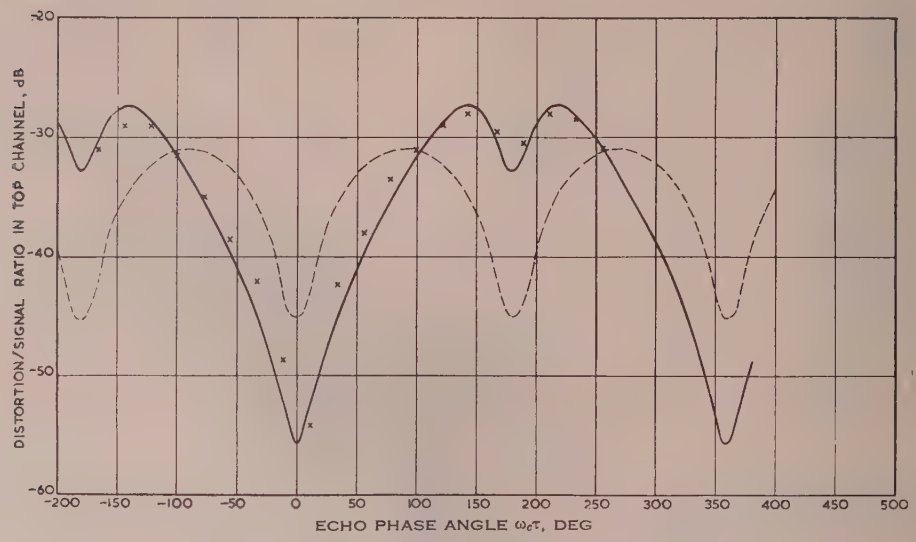


Fig. 3.—Variation of distortion with phase, for a −10 dB echo.
Base band = 0·06 − 1·052 Mc/s; echo delay = 0·083 microsec; peak deviation = ± 2·3 Mc/s.
× Measured by present authors.
———— Computed from eqn. (9).
— — — Computed from small-echo formula.

in the top channel, where

$$F(r) = \frac{r(1 - r^2)\sin \omega_c\tau}{(1 + 2r\cos \omega_c\tau + r^2)^2}$$

and

$$\cos \omega_c\tau = \frac{1 + r^2 - \sqrt{[(1 + r^2)^2 + 32r^2]}}{4r}$$

Figs. 2(*a*) and 2(*b*) show the corrections, as functions of the echo amplitude, that have to be applied to the small-echo formulae, eqns. (4) and (3), when the echo is phased for minimum and maximum distortion respectively. It is seen that, except when the echo amplitude approaches that of the carrier, the correction to the minimum distortion is considerably greater than that to the maximum distortion. The correction to the minimum distortion does not, in fact, become negligible until the echo amplitude is as low as −30 dB.

Fig. 3* shows a comparison between distortion levels in the top channel, measured over a range of echo phase-angles, and values computed from eqn. (9) and from the formula for small echo amplitude, the echo amplitude being −10 dB. The measurements are seen to be in good agreement with the curve derived from eqn. (9) and to depart substantially from Albersheim and Schafer's small-echo formula.

The use of eqn. (10) reduces the discrepancy between the theory and the measurements in Reference 1 to about 3 dB, as shown by the straight portion of the minimum-distortion curve in Fig. 4 [the turn-over at the large-distortion end is computed from the complete eqn. (14)]. It should be pointed out that measurements made by the present authors [see Fig. 8(*b*)] for the same echo amplitude and worst phase are in closer agreement with theory.

What has been said so far applies in the region of small delay time. When the delay time is large the proportionality of distortion to echo amplitude appearing in eqn. (8) has been found by experiment to hold very closely up to amplitudes of −10 dB [see, for example, Figs. 8(*a*) and 8(*b*)]. Measurements made by Albersheim and Schafer (see Fig. 3 of Reference 1) would suggest that the proportionality continues to hold even up to echo amplitudes of 0 dB under long delay conditions. However, it is readily shown (Section 10.2) that when the echo and main-signal amplitudes are equal the frequency-modulation distortion must be zero, whatever the echo delay time. In the special case of small delays this follows immediately from the form of eqn. (9). Such large echoes will generate a very high level of amplitude modulation, which might be expected to break through the limiter to a certain extent and hence to be detected. This may explain the non-zero distortion levels found in Reference 1 for a long-delayed echo having the same amplitude as the main signal.

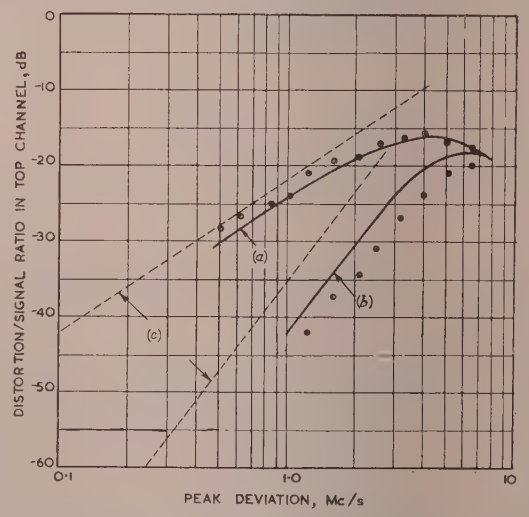


Fig. 4.—Comparison of exact and small-echo formulae, with echo phased for maximum and minimum distortion.
Base band = 0–2 Mc/s; echo delay = 0·2 microsec; echo strength = − 15 dB.
⊙ Measurements by Albersheim and Schafer.
(*a*) Computed from eqn. (13).
(*b*) Computed from eqn. (14).
(*c*) Computed from small-echo short-delay formula.

* In the Figures measurements due to Albersheim and Schafer are plotted as dots and those due to the present authors as crosses, with surrounding boxes where necessary.

5) DISTORTION DUE TO LONG-DELAYED ECHOES WHEN $\omega_D/\hat{\omega}_m$ IS NOT LARGE

The procedure depends on the simple dimensional analysis developed in Reference 10. It is expected on theoretical grounds, and has been confirmed experimentally, that the distortion becomes independent of the delay time when this is sufficiently large. Thus, in this limiting case the ratio of distortion to signal, V_D/V_U, in the top channel is a function of r, ω_D and $\hat{\omega}_m$ only. According to the previous Section, for large delays V_D/V_U is proportional to r, at least up to a -10 dB echo amplitude. Hence $(1/r)(V_D/V_U)$ is a function of ω_D and $\hat{\omega}_m$ only, and since this expression is dimensionless we can write

$$\frac{1}{r}\frac{V_D}{V_U} = \mathrm{F}\left(\frac{\omega_D}{\hat{\omega}_m}\right)$$

where F is a function to be determined. From eqn. (8), when $\omega_D/\hat{\omega}_m \gg 1$,

$$\mathrm{F}\left(\frac{\omega_D}{\hat{\omega}_m}\right) = 5{\cdot}8\left(\frac{\hat{\omega}_m}{\omega_D}\right)^{1{\cdot}5} \exp -2{\cdot}88\left(\frac{\hat{\omega}_m}{\omega_D}\right)^2$$

It will now be shown that there is sufficient experimental information in Reference 1 (largely in Fig. 8) to enable the form of F to be deduced down to $\omega_D/\hat{\omega}_m = 0{\cdot}3$. We consider maximum values of distortion (as a function of echo phase), and suppose that r is so small that V_D/V_U may be considered proportional to r for all delay times [this will certainly be the case for a -20 dB echo, the difference between distortions calculated from eqns. (6) and (11) being only about 0·3 dB]. Then, in the top channel, $(1/r)(V_D/V_U)$ will be a function of τ, $\hat{\omega}_m$ and ω_D. As in Reference 10, these may be combined into two dimensionless groups, which for the present purpose are taken as $\omega_D/\hat{\omega}_m$ and $\omega_D\tau$. $(1/r)(V_D/V_U)$ can thus be completely specified by a series of curves, each corresponding to a fixed value of $\omega_D/\hat{\omega}_m$, plotted against $\omega_D\tau$.

Suppose we attempt to construct such a curve for $\omega_D/\hat{\omega}_m = 0{\cdot}5$, using the measurements in Reference 1 in the manner described in Reference 10. Using Fig. 1 of the latter reference, distortion values must be read off from curves 1, 2, 3 and 4 at abscissa values 0·021, 0·090, 0·250 and 0·4875 respectively. In the first two cases the required points are outside the range of the curves. Values exist, however, from curves 3 and 4, and these are plotted in Fig. 5 of the present paper [for convenience,

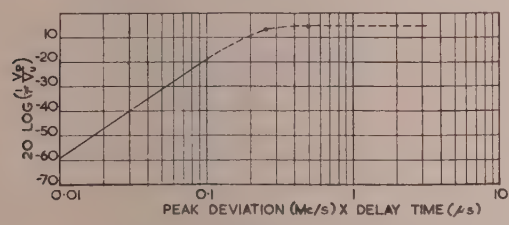


Fig. 5.—Distortion in worst channel, with $\omega_D/\hat{\omega}_m = 0{\cdot}5$ and echo phased for maximum distortion.
○ Scaled from Albersheim and Schafer (experimental).
——— Theoretical (small-delay formula).

the ordinate has been taken as 20 log $(1/r)(V_D/V_U)$ rather than V_D/V_U]. The continuous sloping straight line is plotted from eqn. (6). It is evident that the right-hand point must be in or near the region of long delay, and the distortion curve must be of the form indicated by the broken line. Thus, the limiting value of 20 log $(1/r)(V_D/V_U)$ for large τ is -5, approximately, giving $\mathrm{F}(0{\cdot}5) \simeq 0{\cdot}6$. This is the lowest value of $\omega_D/\hat{\omega}_m$ that can be dealt with in this way, because if the ratio is still further decreased only one point can be deduced from Fig. 1 of

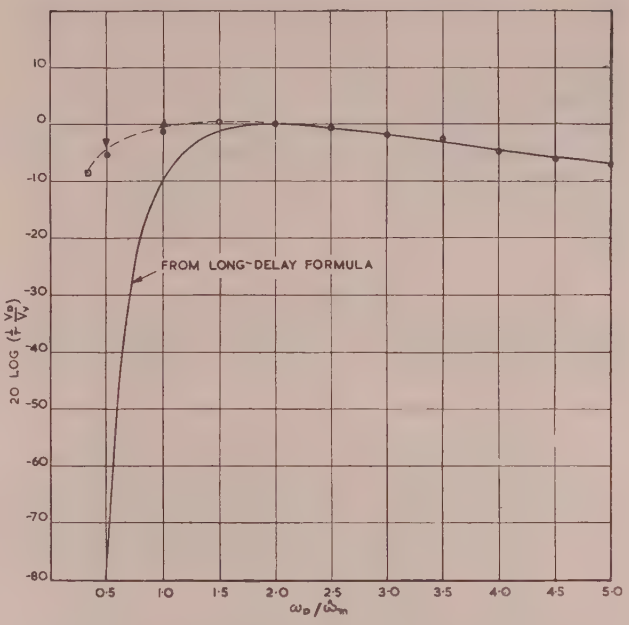


Fig. 6.—Distortion in top channel when delay time is very large.
○ Adapted from Fig. 8 of Albersheim and Schafer.
△ Adopted from Fig. 3 of Albersheim and Schafer.
▽ Adapted from Fig. 5 of Albersheim and Schafer.
□ Derived from curve 4, Fig. 8, of Albersheim and Schafer.

Reference 10, and this does not provide an adequate basis for estimating the limiting long-delay distortion.

Fig. 6 shows 20 log $(1/r)(V_D/V_U)$ plotted against $\omega_D/\hat{\omega}_m$. Most of the points are deduced, in the way just described, from

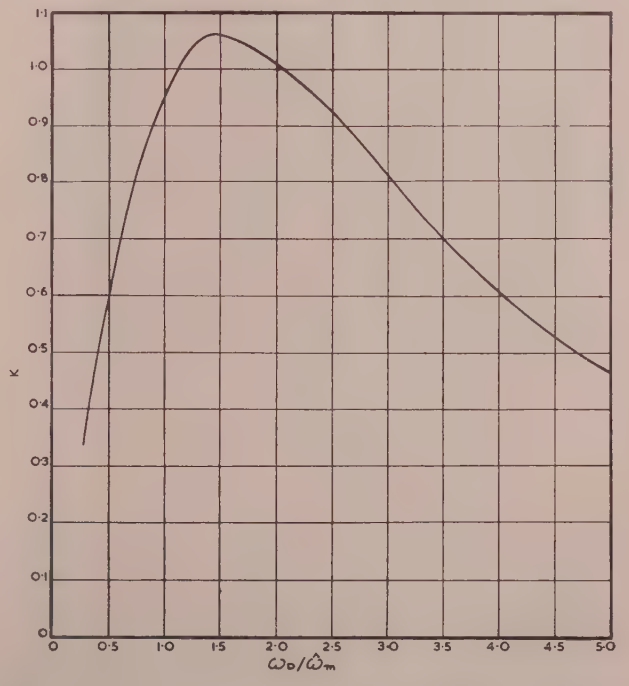


Fig. 7.—Values of K required in eqn. (9) for the distortion due to a long-delayed echo.

Fig. 1 of Reference 10. The derivation of the point for $\omega_D/\hat{\omega}_m = 0{\cdot}3$ will be given in the next Section. When $\omega_D/\hat{\omega}_m = 0{\cdot}5$ and $1{\cdot}0$, additional points can be deduced from two other sets of measurements plotted in Reference 1. It is seen that for $\omega_D/\hat{\omega}_m \geqslant 2$, eqn. (8) agrees closely with the measurements. For lower values of $\omega_D/\hat{\omega}_m$ the distortion calculated from this expression falls off rapidly, while the empirical values decrease only slowly.

It now becomes possible to replace eqn. (8) with an empirical equation in which $\omega_D/\hat{\omega}_m$ is no longer restricted. We may write

$$\text{for } \tau \text{ large,} \qquad V_D/V_U = rK \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where K is a function of $\omega_D/\hat{\omega}_m$ which is plotted in Fig. 7 (derived from Fig. 6).

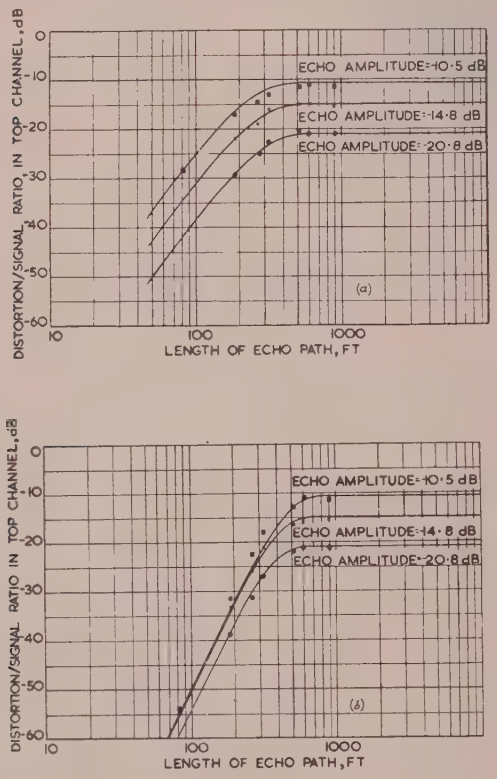


Fig. 8.—Variation of distortion in top channel with echo path length.

Base band = 0·06–1·052 Mc/s; peak deviation = ± 2·3 Mc/s.
Measurements by present authors.

(*a*) Echo phased for maximum distortion.
Curves computed from eqn. (13).

(*b*) Echo phased for minimum distortion.
Curves computed from eqn. (14).

(6) DISTORTION DUE TO ECHOES OF INTERMEDIATE DELAY TIME

It is frequently found that delay times met in practice are neither small enough nor large enough to allow the application of either eqns. (10) and (11) or eqn. (12). This is a situation in which an exact analysis may be expected to present considerable difficulties. Fortunately, it is found that this region is covered with good accuracy by a simple empirical formula, namely

Distortion/signal ratio in top channel

$$= \frac{V_D}{V_U}\left(1 - \exp\frac{-V_{D2}/V_U}{V_D/V_U}\right) \text{ with echo phased for maximum distortion} \quad . \quad (13)$$

$$= \frac{V_D}{V_U}\left(1 - \exp\frac{-V_{D3}/V_U}{V_D/V_U}\right) \text{ with echo phased for minimum distortion} \quad . \quad (14)$$

where V_{D2}/V_U and V_{D3}/V_U are given respectively by eqns. (10) and (11) [or by eqns. (6) and (7) in conjunction with Figs. 2(*a*) and 2(*b*)], and V_D/V_U is given by eqn. (12). Comparison of values calculated from these formulae with measurements covering a range of echo amplitudes is shown in Figs. 8(*a*) and 8(*b*). The agreement is seen to be satisfactory in all cases.

A further check on eqn. (13) over a different range of modulating conditions has been made by comparing it with the measured values plotted in Fig. 8 of Reference 1, as shown in Fig. 9 of the present paper. Adequate agreement is obtained,

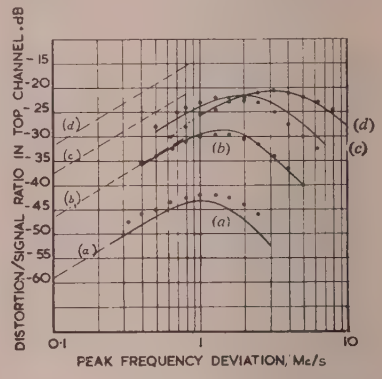


Fig. 9.—Comparison of eqn. (13) with measured values over a range of values of ω_D and $\hat{\omega}_m$, with echo phased for maximum distortion.

Echo strength = − 21 dB; echo delay = 0·5 microsec.
● Measured by Albersheim and Schafer.
— — — Asymptotes computed by Albersheim and Schafer.
———— Calculated from eqn. (13).

(*a*) $\hat{\omega}_m$ = 84 kc/s.
(*b*) $\hat{\omega}_m$ = 360 kc/s.
(*c*) $\hat{\omega}_m$ = 1 000 kc/s.
(*d*) $\hat{\omega}_m$ = 1 950 Mc/s.

particularly for the broader base-bands. Albersheim and Schafer called attention to the discrepancy between their theoretical short delay-time asymptotes (shown as broken lines in Fig. 9) and the measured distortions for small deviation in the case of the larger values of $\hat{\omega}_m$. The use of eqn. (13) removes this discrepancy.

The experimental points associated with curve (*d*) in Fig. 9 can be used to provide an additional point at the left-hand end of Fig. 6. The distortion/signal ratio at the lowest deviation (i.e. 0·634 Mc/s) is −30 dB. Since the echo amplitude is −21 dB, we have, from eqn. (13),

$$K(1 - \varepsilon^{-1{\cdot}708K}) = 0{\cdot}3548$$

giving $K \simeq 0{\cdot}36$, i.e. −8·9 dB. $\omega_D/\hat{\omega}_m$ is 0·325, so that the extreme left-hand point of Fig. 6 follows.

(7) RESULTS DUE TO BENNETT, CURTIS AND RICE

Since the completion of the above work there has appeared a paper in which the same problem has been treated by rather different methods. This paper contains an extensive theoretical treatment of the case of small echo amplitude, and curves are given relating to the phase condition producing maximum dis-

tortion. While these results overlap to some extent those of the present paper, the two papers are largely complementary, since Bennett and his associates do not consider the modification required when the echo amplitude cannot be considered very small, and they do not give numerical results for the phase condition associated with minimum distortion. Even where there is overlapping, the present formulae may be preferred for practical purposes, since numerical integration of the exact formulae in Reference 11, in the region where approximate expressions are not available, involves extensive computation.

Where comparison can be made between the two sets of results, agreement seems to be adequate for practical purposes. As an example, consider the case of 150 ft of air-spaced feeder, both the terminating standing-wave ratios being 1·05 and the echo being phased for maximum distortion. Taking the highest modulating frequency as 1·052 Mc/s and the r.m.s. deviation as 0·82 Mc/s (corresponding to a 240-channel system), we are in a situation where neither the short-delay formula [eqn. (6)] nor the long delay formula [eqn. (8)] applies. From eqn. (13) of the present paper the distortion/signal ratio in the top channel is computed to be −66·9 dB, while from Fig. 5.7 of Reference 11 a value of −68·6 dB is read. The discrepancy, 1·7 dB, is of the order of probable experimental error.

(8) CONCLUSIONS

An important source of intermodulation distortion in the f.m. transmission of frequency-division multiplex is to be found in small echoes generated, for example, by mismatched aerial feeders. The magnitude of this distortion depends on the phase of the echo relative to that of the carrier (except when the echo delay is very large), on the modulation conditions and on the echo amplitude and delay time. Formulae were presented in a previous paper covering distortion levels in certain limiting cases. It now appears that, by the use of eqns. (13) and (14) of the present paper, distortion in the top (usually the worst) channel when the echo is phased for either maximum or minimum distortion can be predicted for most conditions of modulation and echo parameters likely to be encountered.

(9) REFERENCES

(1) ALBERSHEIM, W. J., and SCHAFER, J. P.: "Echo Distortion in the FM Transmission of Frequency-Division Multiplex," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 316.

(2) CARSON, J. R., and FRY, T. C.: "Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency-Modulation," *Bell System Technical Journal*, 1937, **46**, p. 513.

(3) VAN DER POL, B.: "The Fundamental Principles of Frequency Modulation," *Journal I.E.E.*, 1946, **93**, Part III, p. 153.

(4) FRANTZ, W. J.: "The Transmission of a Frequency-Modulated Wave Through a Network," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 114.

(5) GLADWIN, A. S.: "The Distortion of Frequency-Modulated Waves by Transmission Networks," *ibid.*, 1947, **35**, p. 1436.

(6) GOLD, B.: "The Solution of Steady-State Problems in FM," *ibid.*, 1949, **37**, p. 1264.

(7) HUPERT, J. J.: "Normalized Phase and Gain Derivatives as an Aid in Evaluation of FM Distortion," *ibid.*, 1954, **42**, p. 438.

(8) MEDHURST, R. G.: "Harmonic Distortion of Frequency-Modulated Waves by Linear Networks," *Proceedings I.E.E.*, Paper No. 1650 R, May, 1954 (**101**, Part III, p. 171).

(9) HOLBROOK, B. D., and DIXON, J. T.: "Load Rating Theory for Multi-Channel Amplifiers," *Bell System Technical Journal*, 1939, **18**, p. 624.

(10) MEDHURST, R. G.: "Echo Distortion in the FM Transmission of Frequency-Division Multiplex," *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1520.

(11) BENNETT, W. R., CURTIS, H. E., and RICE, S. O.: "Interchannel Interference in FM and PM Systems," *Bell System Technical Journal*, 1955, **34**, p. 601.

(10) APPENDICES

(10.1) Distortion due to Echoes of Short Delay when the Echo Amplitude cannot be considered "Very Small"

The phase distortion due to the echo is given by

$$\phi_D = -\arctan\left[\frac{r\sin(\phi_1-\phi_2)}{1+r\cos(\phi_1-\phi_2)}\right]$$

where ϕ_1 and ϕ_2 are the total instantaneous phase angles of the signal and echo respectively.

If $\omega_M\tau \ll 1$ it can be shown [Reference 1, eqn. (25)] that

$$\phi_1-\phi_2 \simeq \omega_c\tau+\omega_M\tau$$

Thus $$\phi_D = -\arctan\left[\frac{r\sin(\omega_c\tau+\omega_M\tau)}{1+r\cos(\omega_c\tau+\omega_M\tau)}\right]$$

Since $\omega_M\tau$ is small, this can be expanded by Maclaurin's theorem, the result, up to third order in $\omega_M\tau$, being

$$\phi_D = -\arctan\left(\frac{r\sin\omega_c\tau}{1+r\cos\omega_c\tau}\right)-\omega_M\tau\frac{r(r+\cos\omega_c\tau)}{1+r^2+2r\cos\omega_c\tau}$$

$$+\frac{1}{2}(\omega_M\tau)^2\frac{r(1-r^2)\sin\omega_c\tau}{(1+r^2+2r\cos\omega_c\tau)^2}$$

$$+\frac{1}{6}(\omega_M\tau)^3\frac{r(1-r^2)(\cos\omega_c\tau+4r-2r\cos^2\omega_c\tau+r^2\cos\omega_c\tau)}{(1+r^2+2r\cos\omega_c\tau)^3}$$

The term in $\omega_M\tau$ merely alters the base-band amplitude without introducing distortion. When τ is small enough for the present analysis to apply, this term may be neglected. Then, proceeding as in Appendix 1 of Reference 1, the distortion/signal ratio is immediately obtained in the form given as eqn. (9).

(10.2) Proof that Echoes of the Same Amplitude as the Main Signal Produce no F.M. Intermodulation Distortion

If the signal voltage is written in the form $\cos(\omega_c t+\phi_{Mt})$, the addition of an echo of the same amplitude produces the resultant voltage

$$\cos(\omega_c t+\phi_{Mt})+\cos[\omega_c t-\omega_c\tau+\phi_{M(t-\tau)}]$$

which can be rewritten as

$$2\cos\tfrac{1}{2}[\omega_c\tau+\phi_{Mt}-\phi_{M(t-\tau)}]\cos\left\{\omega_c t-\tfrac{1}{2}\omega_c\tau+\tfrac{1}{2}[\phi_{Mt}+\phi_{M(t-\tau)}]\right\}$$

This represents a wave of frequency ω_c, amplitude modulated by the term $2\cos\frac{1}{2}[\omega_c\tau+\phi_{Mt}-\phi_{M(t-\tau)}]$ and phase modulated by $\frac{1}{2}[\phi_{Mt}+\phi_{M(t-\tau)}]$. It is evident that no new frequencies are introduced into the phase modulation, and hence no intermodulation distortion is generated.

Putting $\phi_{Mt} = a\sum_{\omega_m}^{\hat{\omega}_m}\frac{1}{\omega_m}\cos(\omega_m t+\phi_{\omega_m})$

the modified phase modulation is

$$\frac{1}{2}\left[a\sum_{\hat{\omega}_m}^{\hat{\omega}_m}\frac{1}{\omega_m}\cos(\omega_m t+\phi_{\omega_m})+a\sum_{\hat{\omega}_m}^{\hat{\omega}_m}\frac{1}{\omega_m}\cos(\omega_m t-\omega_m\tau+\phi_{\omega_m})\right]$$

$$=a\sum_{\hat{\omega}_m}^{\hat{\omega}_m}\frac{1}{\omega_m}\cos\frac{\omega_m\tau}{2}\cos\left(\omega_m t-\frac{\omega_m\tau}{2}+\phi_{\omega_m}\right) \quad . \quad (15)$$

which is a copy of the original modulation in which the individual tones have been subjected to phase-shifts and amplitude changes dependent on their frequencies.

It is interesting to notice that this provides a physical interpretation of the result—proved in Reference 8 by considering the spectrum of the modulated wave—that a transmission network having a linear phase characteristic and a sinusoidal amplitude characteristic generates no intermodulation distortion. We can generalize the expression for the echo by introducing an additional constant phase angle θ [when the amplitude of the representative tone shown in eqn. (15) would be $\frac{a}{\omega_m}\cos\left(\frac{\omega_m t}{2}-\frac{\theta}{2}\right)$], so that the sum of the unmodulated carrier and echo would be

$$\cos\omega t+\cos(\omega t-\omega\tau+\theta)$$

$$=2\cos\tfrac{1}{2}(\omega\tau-\theta)\cos[\omega t-\tfrac{1}{2}(\omega\tau-\theta)]$$

where ω is general frequency.

Thus, the addition of the echo produces a result which is equivalent to that due to transmission of the modulated carrier through a network having the linear phase characteristic $-\frac{1}{2}(\omega\tau-\theta)$ and the sinusoidal amplitude characteristic $2\cos\frac{1}{2}(\omega\tau-\theta)$. Since the echo has been shown to generate no intermodulation distortion, the required result follows.

621.396.963 : 628.972

Paper No. 2023 R
Mar. 1956

A METHOD OF INCREASING THE AMBIENT ILLUMINATION OF RADAR OPERATIONS ROOMS WITHOUT REDUCTION OF SIGNAL DETECTION THRESHOLD

By C. R. BARNARD.

(*The paper was first received 29th March, in revised form 7th September, and in final form 20th December*, 1955.)

SUMMARY

It is possible to increase the level of ambient illumination in rooms where p.p.i.-type radar displays are used, without any apparent reduction of signal detection threshold, by the use of a narrow-band optical filter superimposed on the face of the cathode-ray tube, and the use of an illuminant complementary to the response of the tube.

Experiments at the Royal Aircraft Establishment, Farnborough, have shown that it is possible, by the use of a number of illuminants lying outside the pass-band of the filter, to create an impression of white ambient lighting in which the colour discrimination of the operations room staff appears to be unimpaired.

It is believed that this development will be of considerable use where it is necessary to display in the same room radar data and plotting-table information. It is also expected that an increase in the level of ambient illumination will lead to a reduction in operator fatigue.

(1) INTRODUCTION

A large number of experiments have been carried out in the past on the visibility of echoes on radar p.p.i. displays under various conditions of ambient illumination. These experiments have been almost entirely restricted to the visibility of persistent-type tubes with incident white light of various levels. An exception to this is the work of Williams and Hanes,* who show that there is a significant effect of colour of ambient light at high intensity. However, their results were obtained with a tube having a blue flash and yellow afterglow—a complex condition involving two colours which tends to confuse the results obtained compared with a monochrome display.

At the Ministry of Supply Station, Defford, radar consoles were installed in a room in which the windows were covered with blue filters, this being the complementary colour to the orange response of the cathode-ray tube used in the display unit (Fig. 1).

The type of installation described in the paper carries the above work a stage further in that it substitutes for blue lighting a form of synthetic white lighting which is orange deficient and does not therefore impair the visibility of the radar display.

(2) PHYSICAL PRINCIPLES

While it is difficult to explain the operation of the system in terms of physiological optics, the underlying physical principles can be clearly seen. If we replace the operator's eye by some physical detector of light rays, such as a photocell, we can assess the performance of the system by assigning certain qualities to the detecting device. Let the detector have a constant sensitivity over the whole range of the visible spectrum and be unable to discriminate between colours. It will therefore be able to distinguish objects only by brightness contrast. Such a device will be quite capable of detecting a target response on a cathode-ray tube display in a room where the ambient light is very dim.

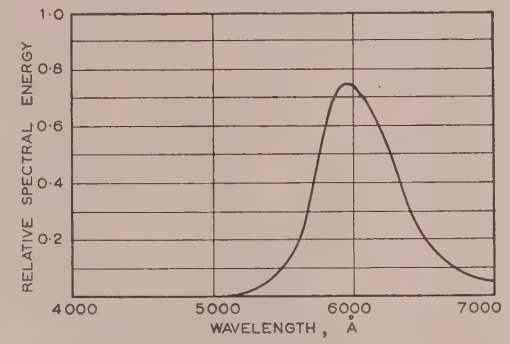


Fig. 1.—Energy distribution of CV429 cathode-ray-tube phosphor [tetr. (Zn : Mg)F_2 : Mn(1)].

As the ambient level is gradually increased the response will be seen as a brightness differential $\Delta B/B$, where B is the brightness of the tube face due to reflected ambient light, and $B + \Delta B$ is the brightness of the spot. As the ambient level is raised further the brightness differential will become too small to be seen by the detector.

If now a filter is added, and has a transmission factor K_1 at the wavelength of the cathode-ray-tube response and K_2 at the wavelength of the ambient illumination, the brightness differential will become $K_1\Delta B/K_2^2B$, the factor K_2 being squared because the ambient light reflected from the tube face will have passed through the filter twice before reaching the detector. If $K_1 > 0{\cdot}9$ and $K_2 < 0{\cdot}1$, the ambient level can clearly be increased by a factor of 90 without reducing the brightness differential.

Although this is a simplification of the more complicated physiological mechanism involved, experiments show that an increase of ambient illumination of this order can be achieved.

The system described above is not dissimilar to the use of a neutral filter, of transmission factor K, for increasing the contrast of television pictures where the brightness differential becomes $K\Delta B/K^2B$. However, with a neutral filter it is not possible to obtain an increase in contrast without incurring an equal decrease in absolute brightness.

Various methods of lighting have been used at the Royal Aircraft Establishment. In the first experiments a stage-lighting batten was used with green and purple filters. This was later replaced by a system using groups of three coloured fluorescent lamps as used for stage lighting. With this system and by the use of dimming circuits it was possible to obtain any colour of ambient lighting lying within the colour triangle shown chain dotted on the trichromatic chart* in Fig. 2.

Finally it was decided, in the interests of economy, to dispense with the complications of a 3-colour system with its dimming

* Williams, S. B., and Hanes, R. M.: "Visibility on Cathode-Ray-Tube Screens: Intensity and Colour of Ambient Illumination," *Journal of Psychology*, 1949, 27, p. 231.

* Staff of the Color Measurement Laboratory, Massachusetts Institute of Technology: "Handbook of Colorimetry" (The Technological Press, M.I.T.).
Wright, W. D.: "The Measurement of Colour" (Adam Hilger, London, 1944).
Chamberlin, G. J.: "The C.I.E. International Colour System Explained" (Tintometer, Salisbury, 1951).

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Barnard is at the Royal Aircraft Establishment.

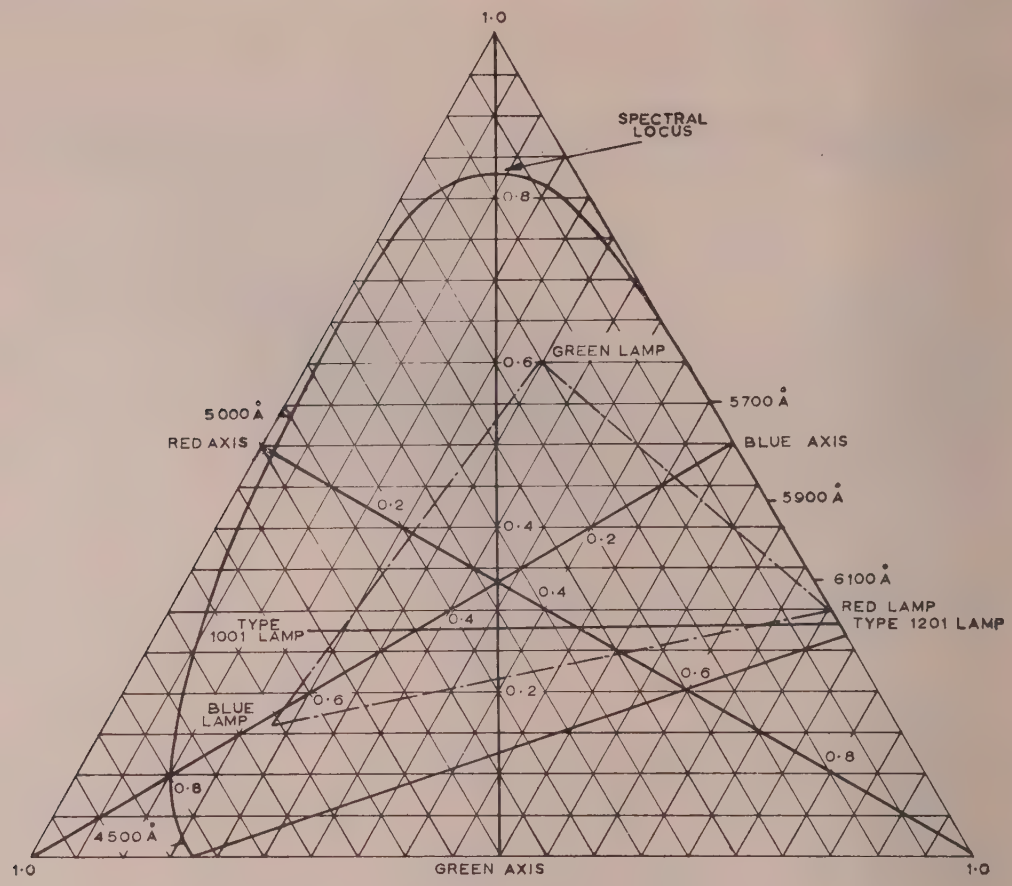


Fig. 2.—Trichromatic plot of room lighting.

circuits and to use two fluorescent lamps specially developed for the project, the blue-green type 1001 and the red type 1201 lamps. The resultant colour obtainable with this system lies along the line joining the colour co-ordinates of the individual tubes, as shown in Fig. 2, its position being dependent on the relative levels of the light sources. With this form of lighting, colour recognition appears unimpaired except for a small number of orange pigments which appear red to the observer.

Several attempts have been made by various organizations to produce a narrow-band-pass optical filter to meet the R.A.E. requirement, and Fig. 3 shows the transmission characteristics of one such filter which was developed for the Ministry of Supply and adopted for the system. Fig. 4 shows the distribution of the ambient lighting with the 2-way transmission curve of the filter superimposed. It will be seen that most of the lighting distribution lies outside the pass band of the filter.

(3) PHYSIOLOGICAL CONSIDERATION

When the human eye replaces the detector it becomes more difficult to assess the performance of the system. In the first place the sensitivity now varies with wavelength, and the sensitivity/wavelength curve is a function of intensity (Fig. 5). Secondly, at intensity levels where photopic (high-intensity) vision occurs, the foveal region of the retina is able to distinguish objects by colour contrast where there is no brightness contrast, e.g. an orange spot on a blue background. Thirdly, stimulation of the retina by light of one colour may affect its sensitivity to

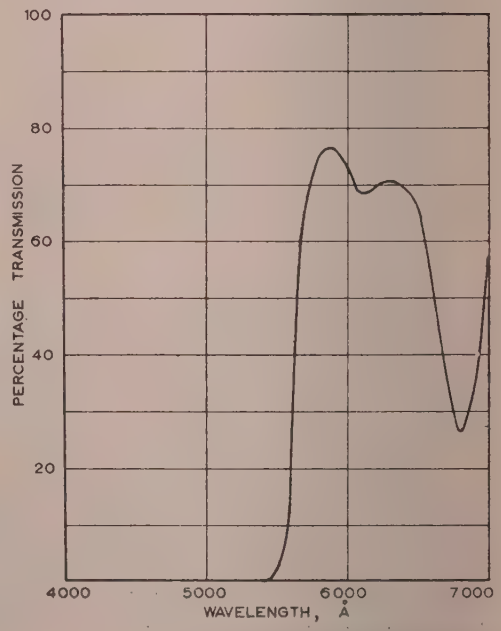


Fig. 3.—Transmission of cathode-ray-tube filter.

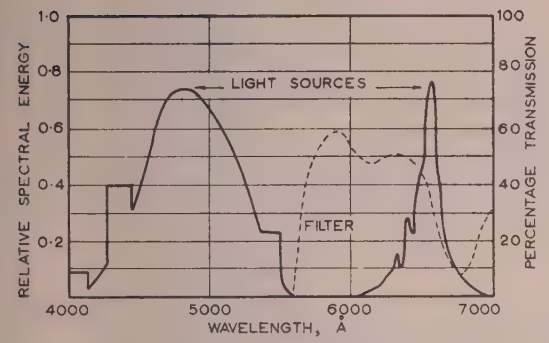


Fig. 4.—Energy distribution of ambient lighting, with transmission curve of cathode-ray-tube filter.

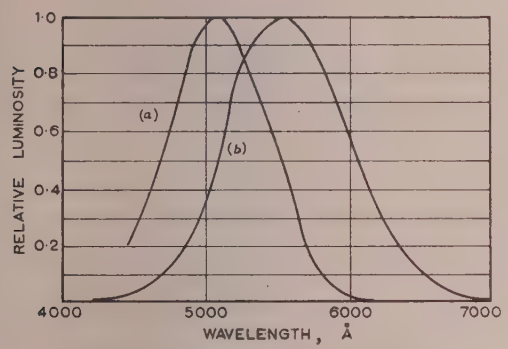


Fig. 5.—Luminosity curves of the eye.
(*a*) Scotopic (low intensity).
(*b*) Photopic (high intensity).

another; this effect of differential adaptation may well be affecting the results obtained. Finally, there is no way of measuring directly the response at the retina for a given stimulus.

All these effects have made it impossible to make direct measurements of the efficiency of the system, but enough is known of the mechanism to make a qualitative assessment. Fig. 1 shows the response curve of the cathode-ray-tube phosphor, and Fig. 6 shows this curve successively modified by the filter and the photopic sensitivity curve. It will be clear from a comparison of Figs. 4 and 6 that, whilst the cathode-ray-tube response is virtually unaffected by this filter, the visual sensation derived from the ambient lighting reflected from the tube face will be extremely small.

(4) EXPERIMENTAL INSTALLATION AND TRIALS RESULTS

The system has been installed in a radar track telling room, where a number of radar consoles are installed with large plotting maps. The level of ambient lighting achieved varies from 7–8 lumens/ft^2 on the plotting tables to 1–2 lumens/ft^2 on the consoles. Great care has been taken to avoid reflections on the cathode-ray-tube masks by painting one wall a dark matt blue and arranging for the consoles to be installed facing this wall. The performance of these displays was checked against identical consoles installed in an unlighted room as a control agency. On a large number of flights the targets were followed to extreme range and subsequently picked up on the return flight. Comparison of the ranges at which signals were lost and detected with the performance of the control installation showed no difference

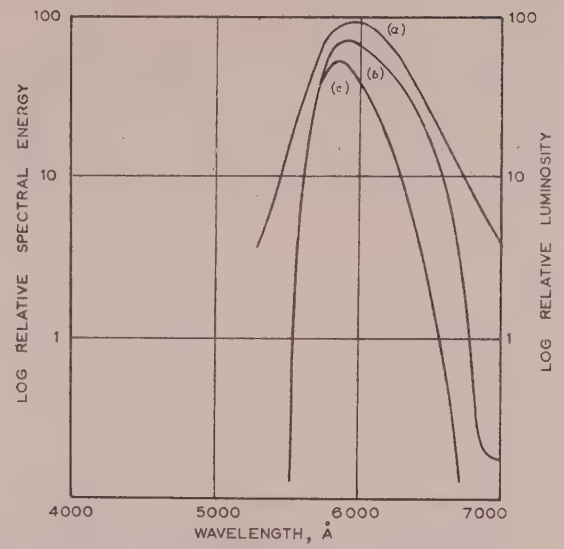


Fig. 6.—Effect of filter and photopic factor on cathode-ray-tube response.
(*a*) Cathode-ray-tube response.
(*b*) Filtered response.
(*c*) Resultant luminosity.

of detection threshold. Thus it follows that use of the "white" illuminant does not impair the radar operator's efficiency.

A number of interesting physiological points have arisen during investigations of the system. Contrary to expectations, the removal of the filter while the operator is tracking a target does not result in the loss of the target. It is possible that this is because any target which holds the attention of the operator is inevitably focused on the fovea, with the concomitant effect of viewing by colour contrast. This would mean that without the filter a new target appearing on the peripheral region of the operator's retina might not be detected. While such a hypothesis is difficult to substantiate by an experiment, there is some confirmation in the operator's instinctive feeling that he is suffering a loss of efficiency under these conditions.

Attempts to measure the signal detection threshold on a noisy tube in conditions of high ambient illumination and in darkness, by injecting synthetic target pulses into the i.f. circuits of the radar equipment, showed no significant difference, the variation being masked by the normal scatter of such measurements. It was therefore decided to attempt to measure the brightness differential threshold of an orange spot on an evenly illuminated orange background as opposed to a noisy background, under the same conditions of ambient illumination as before. While the apparatus was insufficiently sensitive to show any significant difference in the measurement of differential threshold, it was noted that the absolute brightness threshold appeared to be lower in the condition of high ambient illumination than in complete darkness. Here again it is possible only to put forward the conjecture that the lack of any visual reference in the blacked-out room may have made it impossible for the operator to focus accurately on the test object.

(5) CONCLUSIONS

The trials showed that, in this installation, operators plotting on map tables have good colour discrimination, the radar operators do not suffer any loss of efficiency, and the radar supervisors are able to move about freely comparing the plotting-table information with the radar displays. The system has since been

used operationally with apparent success. It has, however, been impossible to obtain any quantitative measurements of the system performance, and it has not been possible to determine whether there has been any reduction in fatigue, although the operators have, after becoming accustomed to the system, expressed a strong preference for working in the lighted room. It is felt that further measurements aimed at determining the system efficiency must be carried out by physiologists rather than by electrical engineers.

(6) ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of the Department of Scientific and Industrial Research, the R.A.F. Radar Research Unit, the Institute of Aviation Medicine, Thorn Electrical Industries Ltd., The British Thomson-Houston Co. Ltd. and Kodak Ltd.

The author is indebted to the Chief Scientist, the Ministry of Supply, and the Controller of H.M. Stationery Office for permission to publish the paper.

DISCUSSION ON 'A STUDY OF SOME OF THE PROPERTIES OF MATERIALS AFFECTING VALVE RELIABILITY'*

SOUTH-WEST SCOTLAND SUB-CENTRE, AT GLASGOW, 2ND FEBRUARY, 1955

SOUTH MIDLAND RADIO GROUP, AT BIRMINGHAM, 28TH FEBRUARY, 1955

Mr. D. S. Gordon (*at Glasgow*): It is apparent that there are different requirements for reliability depending on the purpose to which valves are being put, and while it was true that the Services compromised on civilian types, the converse is now more likely to be the case. For many Service purposes the requirements include a relatively small envelope, good resistance to vibration and shock and a reasonable expectation of life. In general, the industry may be congratulated on developing the miniature all-glass valve to satisfy these requirements generally. However, for many purposes, the larger valve envelopes with separate seal and pin construction may have greater reliability. For teaching purposes, where the valve is repeatedly inserted and removed from its holder, the mechanical strength of the separate base is preferable to the glass base, which tends to crack even when alignment tools are regularly used. With laboratory and industrial equipments using only a few valves, there is generally little advantage in miniaturization, and, in fact, the converse may well be true. Furthermore, it is probably easier for those unskilled in electronic maintenance to replace, without damage to valve or holder, an octal-base valve rather than a 7-pin miniature one. Also, if the same mechanical tolerances are applied, the electrical characteristics of the larger valve can be made more constant than those of the miniature one, and presumably the greater area of emission surface and the possible lower temperatures of the electrodes and envelope will give greater reliability under normal laboratory and industrial conditions.

While not within the scope of mechanical conditions, it would be valuable to receive some authoritative advice on the effect of variations in heater supply voltage and conservative rating of dissipation of valves on their life. When operating under pulse conditions with anode voltages somewhat greater than the rated value, I would like to know whether the rate of deterioration of the cathode is a function of the pulse length, assuming that the mean dissipation is kept constant and within the rated value.

Dr. D. N. Truscott (*at Birmingham*): I am unhappy about the use of log/log scales, particularly in Fig. 7, where three points are plotted, through which a straight line is drawn. The first point, i.e. 'assembly', is plotted at a time of half a day, during which 5% failed. This corresponds to the top line of Table 2, which shows that this process lasts one day. The process of testing, which also lasts one day according to the Table, is plotted on the other side of the ordinate for one day and has a lower failure rate of 0·6%. The point marked 'holding' presumably corresponds to the third line in the Table, which states that 0·005% valves have failed per day over a period of 120 days. If the argument is correct, that the rate of failure falls off steadily as time goes on, I cannot see that it is justified to represent it by a single point at the arithmetical average of this period. It seems to me that log/log scales have been used to smooth out differences, and the straight line is a presentation of facts which is theoretical rather than factual. Furthermore, I wonder whether there is any justification for drawing a straight line through three points which may represent behaviours under very different circumstances. The same criticism does not apply to Fig. 6, in which all the points plotted are failures under roughly constant conditions, and therefore one might expect them to represent the consequences when a constant set of factors are applied.

Mr. J. G. Bartlett (*at Birmingham*): I notice that, in the improved valve, mica ears are used to locate the electrode assembly in the glass bulb. Any cleavage of the mica under vibration might release gas. I would like information on the possible use of stiff flat metal springs, as we often used in cathode-ray tubes.

* ROBERTS, E. A. O'D.: Paper No. 1645 R, February, 1954 (see 101, Part III, p. 197).

621.396.11

Paper No. 1914 R
Oct. 1955

A STUDY OF IONOSPHERIC PROPAGATION BY MEANS OF GROUND BACK-SCATTER

By E. D. R. SHEARMAN, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 14*th March, and in revised form* 23*rd June*, 1955. *It was published in October*, 1955, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 31*st October*, 1955.)

SUMMARY

The results of a year's observations in southern England of long-range back-scatter are analysed. A pulse transmitter coupled to alternative directional aerials was used, and transmissions were made at noon each day on a number of frequencies between 10 and 27 Mc/s.

The echo patterns observed in winter were simple and were formed by echoes from the ground just beyond the skip distances for one- and two-hop F2-layer propagation. In summer, however, ground echoes returning by way of the Es, E and F1 layers arrived nearly simultaneously and were difficult to distinguish. The marked increase in Es ionization in summer could be seen clearly from the echo patterns.

Examples of fixed frequency range-time recordings ($p't$) of back-scatter are given. On one record echoes can be seen at ranges corresponding to each of the ground reflection points for four-hop F2 propagation between England and Malaya. The application of the $p't$ technique to the direct measurement of maximum usable frequencies is described.

To assess the accuracy of the back-scatter technique of skip-distance measurement, comparisons were made between measured scatter ranges and ranges calculated from vertical-incidence ionosphere measurements. The measured ranges were consistently shorter than those calculated. It is uncertain whether this discrepancy arises from approximations in the theory used for the calculations from vertical-incidence data or from inadequate directivity in the aerials used in the scatter measurements. Comparisons of scatter ranges with direct measurements of m.u.f. over the same path are considered desirable to resolve this uncertainty.

(1) INTRODUCTION

In recent years much interest has been aroused by the possibility of utilizing the phenomenon of long-range back-scatter to study the properties of the ionosphere at a point distant from the observer. In the technique used, pulse transmissions are made with a directional aerial at frequencies in the high-frequency band such as are used for long-distance communication. Energy radiated in the direction of the aerial beam is reflected by the ionosphere and returns to the earth beyond the skip distance. The irregularities present on the earth's surface scatter part of the incident radiation in all directions, some energy returning along the path of the incident energy and arriving back at the transmitter after a further ionospheric reflection. This returned energy can be detected in the form of weak echoes of the transmitted pulse and, by measurement of the time delay of the first echo to arrive, the skip distance in the direction of the aerial beam may be deduced.

With suitable equipment and correct interpretation of the results, this technique should provide a powerful tool for the investigation of many problems in short-wave propagation and ionosphere physics. The particular advantage of the method over those used in the past is that the behaviour of waves reflected from the ionosphere at oblique incidence may be studied without the necessity for the co-operation of a distant receiving station.

The paper does not include a survey of the research on back-scatter published by other workers, nor does it give a theoretical analysis of scatter propagation, since this has been done elsewhere.[1] The present purpose is to discuss the results of observations which have been made at the Radio Research Station at Slough to obtain information about the diurnal and seasonal variations of long-distance back-scatter on a number of frequencies. An account is also given of comparisons between these observations and theoretical calculations of skip distance made from vertical-incidence ionospheric measurements. From these comparisons the different modes of propagation present have been deduced and the accuracy of interpretation of back-scatter observations studied.

In the description of the mode of propagation of back-scatter given above, it has been stated that the scattering sources are irregularities on the earth's surface. This is contrary to Eckersley's conclusion[2] that the sources lay in the E region of the ionosphere, but is supported by experimental evidence given by several workers in recent years and by evidence given by the author in another paper.[1] Accordingly, the results described here have been analysed by methods developed in that paper and based on the assumption that the scattering takes place at the earth's surface.

(2) EQUIPMENT

In the experiments described in the earlier paper, the low power and restricted frequency band limited the ground range of the scatter seen to about 1 300 km under quiet ionospheric conditions. In order to obtain information about propagation over greater distances, and to achieve a better signal/noise ratio for the echoes, a transmitter with a peak-pulse power of 150 kW and a pulse length of 120 microsec was built to cover the frequency band 10–27 Mc/s. The sensitivity of the associated receiver for detection of echoes was governed by the atmospheric noise at the aerial.

(2.1) Aerial System

Using the 150 kW transmitter, single-hop echoes could be received from ranges up to 3 000 km, and multi-hop echoes from ranges up to 11 000 km. There was thus a considerable difference in the geographical position of the reflection point for different directions of propagation, and consequently, since the ionization varied with latitude and longitude, there was a marked variation of skip distance with bearing. It was therefore essential to use directional aerials in order to limit observation to one line of shoot by discriminating against echoes returning from directions in which the skip distance had a different value.

Two alternative rhombic aerials were used, the bearings of the main lobes being 80° and 289° E. of N. respectively. Both transmitter and receiver were coupled to the aerial in use, the receiver being connected through a tunable transmit-receive switch. The rhombics, which were of identical dimensions, had an azimuthal beam width of 10° between half-power points at 17 Mc/s. (This beam width is for the squared polar diagram appropriate when using the same aerial for transmission and reception.) The angle of elevation for maximum radiation in the line of shoot was 17° at the same frequency.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

Permanent records were made at intervals by photographing the normal time-base display, and also by taking continuous film records of a brilliance-modulated time-base trace ($p't$ records).

At a later date a rotating horizontal four-element Yagi aerial on a tower 30 ft high was added. This aerial, which was for use with a back-scatter p.p.i. similar to that described by Villard and Peterson,[3] was tuned to 17 Mc/s, had a beam width of 42° for common transmitter-receiver operation, and gave maximum radiation at 23° elevation. The gain of the aerial was 6 dB greater than a dipole.

(2.2) Range Measurement

For convenience, the time bases of the displays were calibrated in kilometres so that the oblique range of the scattering sources from the transmitter could be read off directly. Since this calibration assumed uniform free-space velocity of propagation, the quantity measured was strictly the equivalent free-space path, p', equal to the product of one-half the free-space velocity of light and the time delay of the echo.

The length of the true path followed by a wave travelling to (or from) the ground is shorter than the equivalent free-space path because of the lower velocity of propagation in the layer, but this true path is not normally of interest. In practice, the important quantity is the range along the ground to the scattering sources, and in particular the minimum value of this range, namely the skip distance. This can be found from the equivalent path of the earliest ground-scatter echo by subtracting a small correction which takes into account both the obliquity of path and the non-uniform velocity of propagation. The necessary correction is given in graphical form in another paper;[1] for F2-layer propagation over distances greater than 1 000 km it is of the order of 150 km.

(3) DAILY OBSERVATIONS OF SCATTER

Since August, 1952, the high-power equipment has been used for daily observations at noon on a series of ten frequencies between 10 and 27 Mc/s. The equipment is set up on each frequency and two photographs of the time-base display are taken, the transmitter being connected to the 80° rhombic for one photograph and to the 289° rhombic for the other.

The shape of the echo pattern observed varies considerably throughout the year, the simplest patterns being seen during the winter. On the majority of occasions in winter the F2 layer controls propagation at these frequencies, the ionization density in the E and F1 layers being small and Es ionization being normally absent. The photograph in Fig. 1(*a*) was taken under typical winter daytime conditions and shows the characteristic sharp leading edge and long trailing edge of the normal scatter group from reflection by a single uniform layer.

In summer the ionization in the F2 layer falls and that in the E and F1 layers rises. Under these conditions three simultaneous modes of propagation may be possible at the lower end of the frequency range. In addition, sporadic E ionization, though irregular in position and density, is nearly always present. The scatter patterns then become very complex, as can be seen from Fig. 1(*b*), a photograph taken at noon in July. It is difficult with patterns of this kind to identify the part corresponding to each mode of propagation, since the scatter groups due to different modes overlap on the time base.

(3.1) Analysis of Seasonal Changes in Echo Pattern

The interpretation of such patterns is made easier if the way in which the pattern changes with frequency is studied, and if the characteristics are compared with those expected from

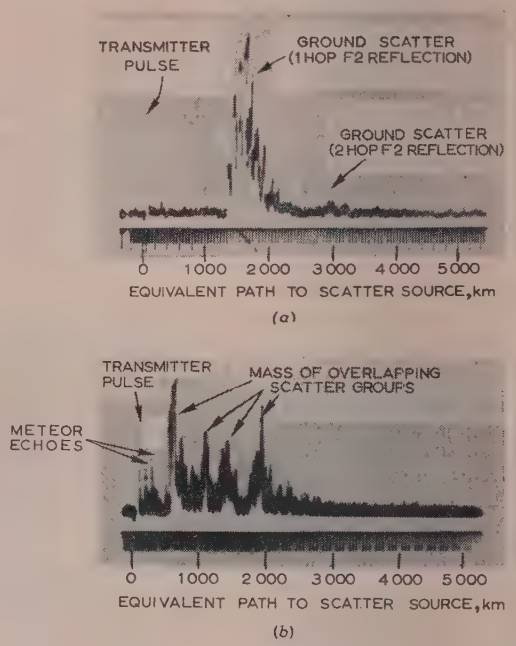


Fig. 1.—Scatter patterns observed on normal time-base display.
(*a*) Winter noon, 13.11.51, 1110 G.M.T. Aerial, rhombic, 80° E. of N. Frequency 18·5 Mc/s.
(*b*) Summer noon, showing complex structure frequently observed, 11.7.53 1050 G.M.T. Aerial, rhombic, 289° E. of N. Frequency 11 Mc/s.

predictions of short-wave propagation conditions such as those compiled by the Radio Research Station. To facilitate comparison, the photographic records of the first year's observations have been analysed and graphs showing the frequency of occurrence of echoes at different oblique ranges (equivalent path-lengths) produced for each month. The time base was divided into 100 km intervals for this purpose, and the presence within each interval of any deflection due to scattered energy was recorded. Three such graphs, for the months of December, March and June, are shown in Fig. 2 and illustrate conditions typical of winter, equinox and summer respectively. No indication of echo amplitude is given, but the effects of the day-to-day changes in range and the systematic variation of range with frequency can be seen. The figure also shows the expected path lengths of the leading edge of ground-scatter echo groups for E, F1 and F2 propagation calculated from predictions of average ionospheric conditions.

(3.2) Winter

Examination of Fig. 2(*a*) shows that scatter was present at some ranges on 100% of the occasions for frequencies up to 19 Mc/s. In order to compare the observed scatter ranges shown in this diagram with the predictions, it should be remembered that all the observed scattered echoes are shown, whereas the prediction is for the leading edge only. However, the average value of the observed leading-edge path length for each frequency must lie between that for the first non-zero ordinate and that for the first 100% ordinate, and should agree with the predicted average value for F2 propagation. The agreement is good for the lower frequencies in December, but becomes progressively poorer as the frequency is increased. If the scatter results are being interpreted correctly the predicted maximum usable frequency (m.u.f.), is about 2 Mc/s low for a range at 3 000 km; this discrepancy is discussed in Section 7.

It can be seen from Fig. 2(*a*) that the E and F1 modes should

give a longer path (and skip distance) than the F2 mode. This implies that the echoes will be received at a low angle of elevation, for which the response of the aerial will be poor, and the received signals consequently weak, in relation to the F2 signals.

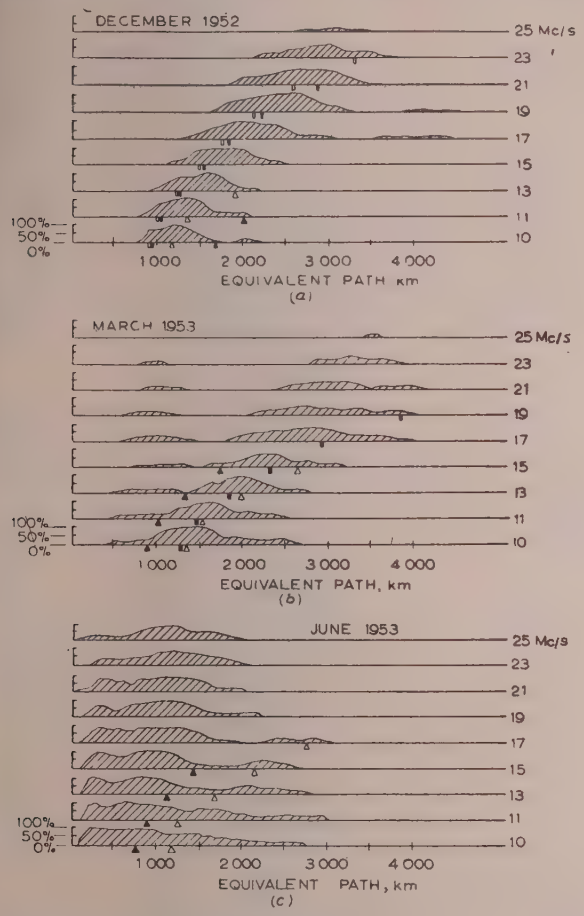


Fig. 2.—Relation between percentage of occasions on which scattered echoes were seen and range of the echoes.

(*a*) December, 1952.
(*b*) March, 1953.
(*c*) June, 1953.

Time, noon. Aerial, rhombic. 80° E of N.
Path length of leading edge from published predictions:

▲E
△ F1
▮ F2
} = For direction of centre of main lobe of aerial.

[F2 = For direction of southern extremity of main lobe.

Fig. 2(*a*) also shows that the limiting range of one-hop scatter which is represented for any frequency by the trailing edge of the first group, increases with frequency up to a maximum value, for some observations, of about 3 600 km. The angle of elevation for this range will be 2° or 3° only, so that the limiting tangential-ray condition is approached at the high frequencies. At the lower frequencies, echoes were not received up to the limiting range. This would be expected because of the poor low-angle response of the aerial system at these frequencies, and because the enhancement of the echo strength by focusing is limited to the neighbourhood of the skip distance.

At frequencies of 17 and 19 Mc/s some echoes were observed on a few occasions at approximately twice the range of the main group. This is an example of scatter from the two-hop focusing area of the kind discussed in Section 5.

(3.3) Equinox

Fig. 2(*b*) shows that in March a main scatter group was observed at a longer range than in December, and, in addition, scatter is seen on about 25% of the occasions at a shorter range than the main group. This range, which is much too short to be explained by propagation by normal layers, indicates[1] that the echoes are due to ground scatter reflected by Es ionization.

The predictions indicate that up to 15 Mc/s scatter reflected by the normal E layer should be seen at a shorter range than the F2 scatter, but its presence cannot be confirmed since the patterns for the two modes overlap. For frequencies above 15 Mc/s, where F2 propagation only is predicted, the discrepancy between prediction and observation is more serious than that in December.

(3.4) Summer

Fig. 2(*c*) shows that the Es mode is predominant in June, echoes being present for nearly 100% of the occasions on the lower frequencies, and for 75% at frequencies even as high as 25 Mc/s. The predictions for this month indicate that F2 propagation should have been impossible in the frequency range 10–25 Mc/s, since a ray at an angle of elevation sufficiently high to penetrate the F1 layer should have been above the escape angle for the F2 layer. This is confirmed by the diagram, which shows scatter at the ranges expected for F1 reflections on 30% of occasions, but no scatter that could be attributed to F2 reflection.

Fig. 2(*c*) also indicates the presence of echoes at ranges as short as 100–500 km. These appear to be due to direct scatter from sources in the E region, as discussed in Section 4.

(4) SCATTER REFLECTED BY Es IONIZATION

It is not possible to analyse the observations of Es-reflected scatter in the same detail as those of scatter reflected by the normal layers, because of the very variable nature of the ionization. However, a few points of interest can be mentioned.

The strength of Es-reflected echoes varied greatly from day to day, and it varied also with the bearing of the aerial used, the amplitude at the receiver input varying from a few microvolts up to a millivolt. Observations with an Adcock direction finder showed that the Es-reflected echoes were at times arriving from a bearing very different from that of the transmitting-aerial beam, while the F2-reflected scatter, seen at the same time, came from approximately the direction of the beam. Part of the variation in signal strength was therefore due to the directivity of the aerial.

Undoubtedly the best method now available for the study of Es-reflected scatter is the p.p.i. technique described by Villard and Peterson[3] and de Bettencourt[4]. Observations with similar equipment at the Radio Research Station are now being made to supplement the fixed-aerial observations.

It should be mentioned that, in addition to the ground scatter reflected by Es ionization, weaker echoes, which appear to be scattered direct from the Es cloud responsible for the reflection, are occasionally received. These echoes appear at half the range of the ground echoes, and occur when the intensity of the ground scatter is high. At the lower frequencies very weak, rapidly fading echoes, at a range corresponding to the part of the E layer illuminated by the main lobe of the aerial, are also seen at times and form a background signal on which transient meteor echoes are superimposed. This second type of direct

echo does not appear to be necessarily associated with Es ionization, and it has been observed when no ground scatter echoes could be seen.

(5) CONTINUOUS RECORDINGS OF SCATTER

Examples of continuous ($p't$) scatter recordings have been presented by Eckersley,[2] Kōno,[5] Dieminger[6] and Hartsfield *et al.*,[7] so that a detailed discussion of such experiments is unnecessary here. However, two records made at Slough are of interest and are shown in Figs. 3A and 3B.

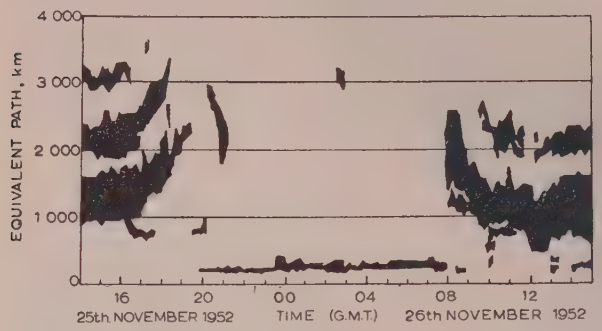


Fig. 3A.—25 h continuous recording of scatter. 10·05 Mc/s. Rhombic aerial directed 289° E. of N.

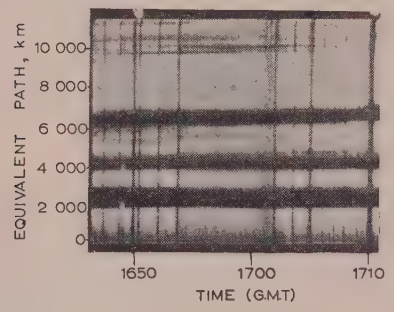


Fig. 3B.—Continuous recording showing 4-hop scatter at 10 000 km range. 5th September, 1952. 17 Mc/s. Rhombic aerial directed 80° E. of N.

The first (Fig. 3A) is a 25-hour record made at a frequency of 10 Mc/s, using an aerial directed towards the west. During the afternoon of the 25th November, 1952, three distinct groups of scatter were recorded at ranges corresponding to the skip focusing areas for three-hop F2 propagation with equal hop-lengths. In the evening, the range of the first group increased until it faded out. The ranges of the leading edges of the second and third groups increased approximately in proportion to that of the first, suggesting that the intermediate ground reflections were specular. A temporary reappearance of scatter at 3 000 km was recorded at 2010 G.M.T. and again at 0230 G.M.T., but during most of the night only the weak direct E scatter referred to in Section 4 was present.

Fig. 3B shows the very long range of the echoes seen at certain times at the higher frequencies. The path length of the fourth group of scatter on this record is 10 000 km, corresponding to the range of Malaya along the great circle through the line of shoot of the aerial. The lengths of the successive hops on this occasion were 2 000, 2 000, 2 000 and 4 000 km. The local time at the reflection point for the 4 000 km hop was 2 300 hours for a local time of 1 700 hours at Slough. A long hop at this frequency (17 Mc/s) might therefore be possible over the dark part of the path together with short hops over the sunlit part.

It is, however, possible that the 4 000 km hop was not due to a single reflection at the layer but to a more complicated trajectory. A number of observations have suggested that such trajectories can occur. For example, on several occasions two groups of scatter have been seen, one group at double the range of the other. This is the condition normally associated with two-hop propagation, for which scatter would be returned from the one- and two-hop skip-focusing zones. However, on these occasions the first group was observed to fade out, leaving only the two-hop scatter group with a range of 4 000–6 000 km. It is possible that this very long trajectory, apparently without an intermediate ground-reflection point, was made possible by two successive reflections from the layer, appropriate tilts being present at the reflection points. Whether this is the correct explanation of the phenomenon is not yet known.

It has also yet to be ascertained whether complex modes of propagation such as this are of practical significance in long-distance high-frequency communication.

(6) MEASUREMENT OF THE UPPER LIMITING FREQUENCY FOR SCATTER RECEPTION

The continuous-recording technique was adopted in an experiment to measure throughout a 24-hour period the maximum frequency at which scatter could be received, for comparison with predicted m.u.f.'s for the same direction.

(6.1) Experimental Procedure

A series of frequencies, spaced at intervals from 10 to 21 Mc/s, was chosen, and transmissions were made on a rhombic aerial directed 80° E. of N. At the beginning of the period (0800 G.M.T., 15th May, 1952) the transmitter frequency was increased in steps, beginning at 10 Mc/s. At 19 Mc/s the range of the scatter was 3 000 km, while at 21 Mc/s no echoes were observed, showing that the maximum frequency for scatter reception lay between the two. Alternate ten-minute recordings were made on the two frequencies until, at 1140 G.M.T., echoes were seen on 21 Mc/s, indicating that the maximum frequency had risen above this value.

The technique of alternate ten- or five-minute recordings on different frequencies was used throughout the 24 hours to establish the fade-out and recovery times at every frequency. The results are presented in Fig. 4, in which the shaded areas represent the limits between which the highest frequency for scatter reception was found to lie throughout the day.

The scatter faded out or recovered on all the frequencies at an

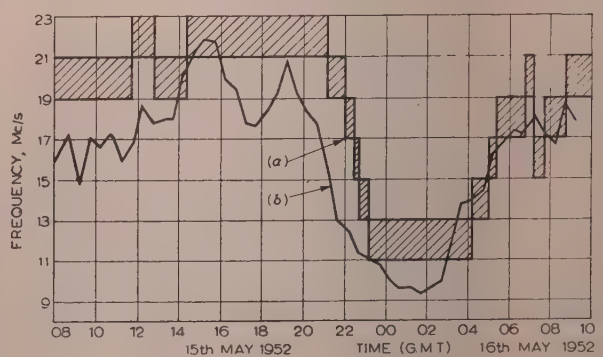


Fig. 4.—Measurement of 3 000 km m.u.f.
(*a*) By back-scatter.
(*b*) From Lindau vertical-incidence measurements.

oblique range of 3 000 km, corresponding to a skip distance of 2 850 km. If it is assumed that the scatter came from the direction of the line of shoot of the rhombic aerial (80° E.), then the scatter limiting-frequency curve should be nearly the same as the m.u.f. curve for propagation over a path of 3 000 km on this bearing.

(6.2) Comparison with Vertical-Incidence Measurements

The 3 000 km m.u.f. curve calculated from vertical-incidence measurement at Lindau, Germany, has been plotted in Fig. 4. Lindau is about half-way between Slough and the reflection point and, in consequence, to allow for the local time difference, the m.u.f. curve has been shifted by 45 min. The agreement between the scatter and Lindau m.u.f. curves is good during the period between 0400 G.M.T. and 0900 G.M.T. on the 16th, but not during the daytime or evening of the 15th. This discrepancy may be due to a genuine difference between the ionosphere at Lindau and the ionosphere at the reflection point, or it may be due to one of the possible sources of error discussed in Section 7.

(6.3) Limiting Range for Single-Hop Propagation

One interesting result of this experiment is that the limiting range for scatter reception was found to be 3 000 km and not 4 000 km, which is normally assumed to be the limiting range for one-hop propagation. In the observations discussed in Section 3 some single-hop scatter was seen at ranges greater than 3 000 km, but such instances were rare. If this lower value for the limiting range applies to one-way propagation it is a matter of considerable importance, since in the planning of long-distance communication at present limiting m.u.f.'s are calculated by means of m.u.f. factors appropriate to hop lengths of 4 000 km. Unfortunately, it is not certain that the results are applicable to one-way communication, because the conditions are different.

The factors governing the limiting range of one-hop propagation when the transmitter and receiver are at opposite ends of the path are the height of the layer and the gain of the aerials at low angles of elevation. For scatter propagation an additional factor is the re-radiating efficiency of the scattering objects at low angles of elevation. With scattering objects such as waves at sea and ground irregularities of low height the re-radiation efficiency might be expected to fall off at low angles (say, $< 5°$), so it is possible that the limiting range for scatter reception is reached before that for one-way transmission.

It is not at present certain, therefore, whether the results of the experiment indicate an upper limit of 3 000 km for one-hop propagation.

(7) COMPARISON OF NOON SCATTER OBSERVATIONS WITH VERTICAL-INCIDENCE MEASUREMENTS

The best method of determining the accuracy of interpretation of scatter results is to compare the m.u.f. deduced from scatter observations for a particular path with a direct measurement of m.u.f. over the same path. The m.u.f. measurement can be made at any time of day if oblique-incidence *p'f* equipment is available, i.e. an ionospheric sounder having a transmitter and receiver at opposite ends of the path.[8,9] Alternatively, field-strength recordings can be made of a transmission from a distant c.w. transmitter on a fixed frequency, and observations made of the times of fade-out and recovery of the received signal.[10]

Unfortunately, no direct measurements of m.u.f. were available to check the values deduced from the noon scatter observations. However, an indirect check has been made by comparing the observed minimum path-length of the scatter with the value calculated from vertical-incidence measurements made in the neighbourhood of the reflecting point at nearly the same time. The observations made at 17 Mc/s during November and December, 1952, gave values for the skip distance in the direction 80° E. varying between 1 250 and 1 850 km, corresponding to a reflection point 625–925 km distant from Slough and in the neighbourhood of the ionospheric station at Lindau.

The minimum path-lengths calculated from the Lindau data are plotted against the scatter path-lengths in Fig. 5. The cal-

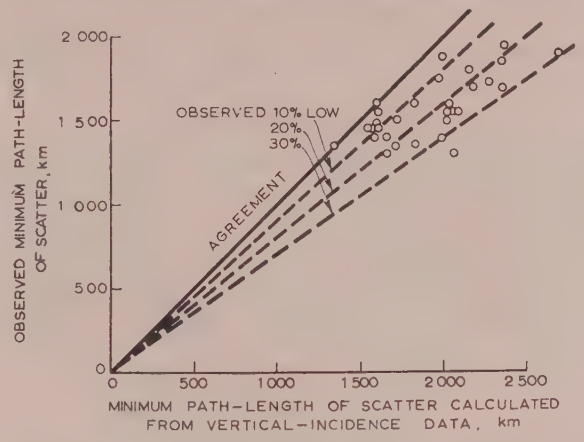


Fig. 5.—Comparison of observed and calculated path-length of leading edge of scatter group. Noon, November–December, 1952.

culated values are for ordinary-ray propagation. The points representing individual observations are scattered, as would be expected from the random changes in F2-layer critical frequency over the area at which reflections were taking place. However, there is also a systematic difference of 10–15% between observed and calculated values, the observed path-lengths being low. This result is similar to that obtained when the same measurements were compared with predictions (Fig. 2) and when a like comparison was made in the experiment discussed in Section 6.

Various possible sources of this discrepancy have been studied, both in the calculation of m.u.f. from vertical-incidence measurements and in the interpretation of the scatter observations.

(7.1) Errors in Calculating Oblique M.U.F. from Vertical-Incidence Measurements

The assumptions made in the theory used to calculate skip distance (or minimum path-length) from vertical-incidence measurements are as follows:

(*a*) The surfaces of constant ionization-density in the ionosphere are spherical and concentric with the earth's surface.

(*b*) Approximate compensation for the increase in F2 skip distance caused by the bending of rays as they pass through the E layer is made by the method used to find the height of maximum ionization for the F2 layer: this height is calculated from the vertical-incidence record by single-layer theory, but with the recorded virtual heights of reflection, which include retardation in the E layer (Appleton and Beynon[11]).

(*c*) At long ranges the skip distance for the extraordinary ray is nearly the same as that for the ordinary ray, which may be calculated from no-field theory (Tremellen and Cox[12]).

From estimates which have been made, it does not appear that either of the first two assumptions can introduce any appreciable error into calculations of skip distance for noon in winter. The effect of the third assumption cannot be ascertained until oblique-incidence *p'f* measurements are made over long distances. The little evidence at present available, however, supports the estimates that the difference between ordinary-

and extraordinary-ray skip distances becomes smaller at long ranges. If this is so, the error introduced by this assumption may also be small.

A further possible explanation for the discrepancy may be that the mode of propagation is not adequately described by the present theory, for reasons not yet understood. That authorities operating long-distance radio circuits have reported consistent use of frequencies higher than those predicted from the theory is suggestive, the differences being too great to be explained by known modes of propagation. This possibility also could be investigated by oblique-incidence $p'f$ experiments.

(7.2) Errors in Interpretation of Scatter Observations

One assumption made in interpreting the scatter patterns must also be considered. It has been assumed that the echoes forming the leading edge of the scatter pattern come from the bearing of the centre of the main lobe of the aerial. An estimate of the effect of off-beam echoes for noon in December, 1952, has been made with the help of the published world contours of critical frequency and m.u.f. factor from which m.u.f. predictions are made. From these charts, Fig. 6 has been constructed to

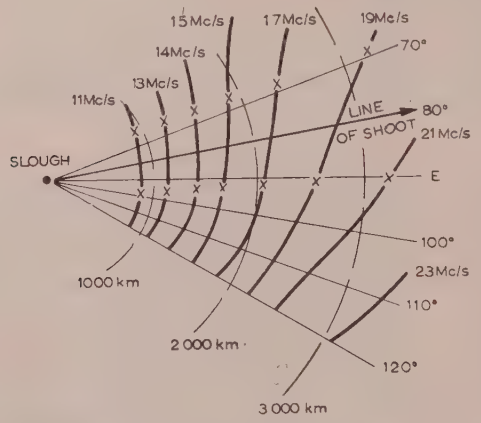


Fig. 6.—Skip-distance contours for noon, December, 1952, calculated from m.u.f. prediction charts.

Each contour shows the skip locus for the frequency indicated.
X—Limits of aerial beam for 20 dB loss.

show the locus of the skip distance at the frequencies used in the noon measurements for bearings from 50° E. to 120° E. It was assumed that echoes could only have been detected if they came from directions in which the aerial response for common transmitter-receiver operation was not more than 20 dB below the maximum. The width of the beam for 20 dB loss has been found from the calculated polar diagram for each frequency, and is plotted on the corresponding contour. It can be seen that the skip distance at the southern limit of the aerial beam is less than that along the line of shoot, the difference increasing with range.

Applying these results to the observations of scatter for December, the calculated values of p'_{min} in Fig. 5 should be reduced by from 5% at 1 500 km to 8% at 2 500 km. This would give a better correlation between observed and calculated values, although there would still be a systematic error of about 10%. Predictions of the minimum path-length at the edge of the aerial beam are plotted in Fig. 2 for comparison with the diagram for December.

The discrepancies seen in the experiment described in Section 6 may also have arisen partly or wholly from the same source. The predicted m.u.f.'s for the month of May appear to be low for the particular day of this experiment and so cannot be taken as a reliable guide, but an examination of the predicted m.u.f. contours shows that errors of 2 Mc/s in m.u.f., or one hour in fade-out time, might have been expected to occur.

(7.3) Evidence for Reception of Off-Beam Scatter

Some experimental evidence which demonstrates the presence of off-beam scatter is of interest. On one occasion the scattered echoes from pulse transmissions from a rhombic aerial were received on an Adcock cathode-ray direction finder. A strobe was arranged so that the echoes in a selected part of the scatter group seen on a normal time-base display could be presented on the bearing display. The pattern seen on the bearing display consisted of a mass of ellipses at slightly different bearings, fading at random. When the strobe was moved through the scatter group a progressive change in the mean bearing was seen, showing that the predominant source of scatter lay in a different direction for each part of the pattern.

It was not possible at this time to carry out a more detailed investigation of these effects with the Adcock direction finder. However, at a later date a rotatable directional aerial became available and provided a very convenient method of studying the effect of off-beam scatter. During one observation, this aerial, a Yagi with a half-power beam width of 40°, was being used for transmission and reception on 17 Mc/s. Scatter reflected by the F2 layer was seen to the east, south and west, with the range increasing gradually from south to east and from south to west. When the aerial was directed towards the south, the path-length of the scatter was shortest and the scatter group had a sharp leading edge. As the aerial was rotated away from the south in either direction, the trailing part of the pattern increased in amplitude and the leading part decreased, so that the sharp leading edge degenerated into a gradual rise. This is the result which would be expected if the scatter from each elementary sector gave a pattern with a sharp leading edge[1] and if at the same time the skip distance was increasing smoothly from south to north. The gradual rise of the leading part of the pattern would then be introduced by the poor response of the aerial to the shorter-range scatter at the southern limit of the beam.

This type of scatter pattern is frequently seen when the rhombic aerials are used, particularly at long ranges, thus lending support to the suggestion that the measured skip distance may be shorter than the calculated value because it is for a direction to the south of that which has been assumed in the calculation.

It should be mentioned that experiments with the rotating aerial indicate that a smooth variation of skip distance with azimuth is comparatively rare. It is, however, doubtful whether the true shape of the skip locus can be determined until an aerial with a much narrower beam is available. The reverse process of calculating the observed patterns from assumed skip loci is, however, possible,[1] and would provide a useful guide to the observer.

(8) CONCLUSIONS

The use of a high-power transmitter in association with directional aerials has made it possible to obtain back-scatter echoes of good signal/noise ratio regularly, and to study the diurnal and seasonal variations in the echo patterns observed at various frequencies. The interpretation of these echo patterns is straightforward in winter but becomes very complex in summer, when energy propagated by way of several ionospheric layers returns to the receiver at nearly the same time. A useful guide to the interpretation of complex patterns is provided by published predictions of radio propagation conditions. From such pre-

dictions it is possible to calculate the expected characteristics of the echo pattern and thus to identify the parts of the pattern produced by the various modes of propagation. This analysis shows clearly the marked increase of sporadic E ionization during the summer months.

Before the back-scatter technique can be used to measure skip distance with confidence, it would be desirable to check the skip distances or m.u.f.'s obtained from scatter observations against direct measurements of the same quantities. Such direct measurements, which could be provided by swept-frequency oblique-incidence pulse measurements, or by field-strength recordings of distant transmitters, were not available for comparison with the observations described in the paper. As an indirect check, however, a comparison was made between the range of scatter and the range expected from vertical-incidence measurements made near the reflection point. The observed range was found to be consistently less than that predicted.

A study of the known approximations in the theory used to calculate scatter ranges from vertical-incidence data appears to discount them as a cause of the discrepancy, although sufficient evidence is not available to indicate whether the effect of the earth's magnetic field is correctly allowed for by the theory. It is, however, possible that the theoretical model used for the calculations gives an inadequate representation of the true mode of propagation for reasons not at present known. In this connection it is of interest that radio communication over long distances on frequencies consistently higher than those considered possible from the theory has been reported.

A further suggested explanation of part or all of the discrepancy is that the shortest-range scatter came from a direction away from the line of shoot of the aerial. An estimate indicates that up to half of the observed error might have been due to this cause. It is considered that the reception of such scatter from the edge of the main beam may seriously limit the accuracy of skip-distance measurement by the scatter technique, and it is therefore desirable that further theoretical analysis be pursued and experiments with aerial systems undertaken to find the magnitude of the errors introduced and the steps necessary to overcome them.

In order to eliminate the indirect check of the measurements by back-scatter of skip-distance and m.u.f. which has been necessary in this work, it is most desirable that direct measurements of these quantities should be available in future experiments. The most convenient method would be to compare a scatter measurement of m.u.f. over a particular path with a direct measurement of m.u.f. over the same path by an oblique-incidence $p'f$ equipment. An experiment of this kind should enable the accuracy of back-scatter measurements to be determined without ambiguity.

(9) ACKNOWLEDGMENT

The work described was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The author wishes to acknowledge the help and advice of Mr. A. F. Wilkins, Mr. W. R. Piggott and Mr. L. T. J. Martin. Mr. C. G. McCue, of the Australian Department of Supply, co-operated in the work for a short period and assisted in the analysis of results.

(10) REFERENCES

(1) SHEARMAN, E. D. R.: "The Technique of Ionospheric Investigation using Ground Back-scatter" (see next page).

(2) ECKERSLEY, T. L.: "Analysis of the Effect of Scattering in Radio Transmission," *Journal I.E.E.*, 1940, **86**, p. 548.

(3) VILLARD, O. G., Jr., and PETERSON, A. M.: "Scatter Sounding: A Technique for the Study of the Ionosphere at a Distance," Transactions of the Institute of Radio Engineers Professional Group on Antennas and Propagation, August, 1952, PGAP-3.

(4) DE BETTENCOURT, J. T.: "Instantaneous Prediction of Ionospheric Transmission Conditions by the Communication Zone Indicator," *ibid.*

(5) KŌNO, T.: "Experimental Study on Scattered Echoes," Report of Ionospheric Research in Japan, 1950, **4**, No. 3.

(6) DIEMINGER, W.: "The Scattering of Radio Waves," *Proceedings of the Physical Society* (B), 1951, **64**, p. 142.

(7) HARTSFIELD, W. L., OSTROW, S. M., and SILBERSTEIN, R.: "Back-Scatter Observations by the Central Radio Propagation Laboratory," *Journal of Research of the National Bureau of Standards*, 1950, **44**, p. 199.

(8) BEYNON, W. J. G.: "Propagation of Radio Waves," *Wireless Engineer*, 1948, **25**, p. 322.

(9) SULZER, P. G., and FERGUSON, E. E.: "Sweep-Frequency Oblique-Incidence Ionosphere Measurements over a 1 150 km Path," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1124.

(10) HARTSFIELD, W. L., and SILBERSTEIN, R.: "A Comparison of C.W. Field Intensity and Backscatter Delay," *ibid.*, 1952, **40**, p. 1700.

(11) APPLETON, E. V., and BEYNON, W. J. G.: "The Application of Ionospheric Data to Radio-Communication Problems: Part I," *Proceedings of the Physical Society*, 1940, **52**, p. 518.

(12) TREMELLEN, K. W., and COX, J. W.: "The Influence of Wave-Propagation on the Planning of Short-Wave Communication," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 200.

[The discussion on the above paper will be found on page 232.]

621.396.11

Paper No. 1915 R
Oct. 1955

THE TECHNIQUE OF IONOSPHERIC INVESTIGATION USING GROUND BACK-SCATTER

By E. D. R. SHEARMAN, B.Sc.(Eng.), Associate Member.

(*The paper was first received 14th March, and in revised form 23rd June, 1955. It was published in October, 1955, and was read before the* RADIO AND TELECOMMUNICATION SECTION *31st October, 1955.*)

SUMMARY

The skip distance in short-wave propagation is measured by observing the time delay of echoes scattered from the earth's surface beyond the skip distance. A theoretical study shows that the method should give good accuracy in determining skip distance, but cannot yield unique values of the height and critical frequency of an ionospheric layer at a distant point.

The radar equation is used to find the intensity of back-scatter from the ground, and an approximate calculation indicates that the irregularities present on land and sea should be sufficient to explain the observed echo strength. The effect of ray focusing by the ionosphere is considered.

Experimental results confirm the conclusions of other workers that the predominant sources of echoes are on the ground and not in the E region. An empirical correction for the effect of the earth's magnetic field shows that no error in the location of the sources is introduced by using no-field theory in the analysis of results.

LIST OF PRINCIPAL SYMBOLS

D = Ground range to scatter source.
D_{skip} = Skip distance.
p' = Equivalent free-space path-length to scatter source.
p'_{min} = Minimum value of p'.
f = Transmitted frequency.
f_v = Equivalent vertical-incidence frequency [eqn. (1)].
f_o = Critical frequency.
f_H = Magneto-ionic gyro frequency.
h'_v, h' = Virtual heights of reflection.
h_E = Height above ground of E layer.
h_{tan}, f_{tan} = Height and frequency co-ordinates of contact point of the tangent drawn to an $h'f$ curve from the origin.
h_m = Height of maximum ionization of a parabolic layer.
y_m = Semi-thickness of a parabolic layer.
$h_o = h_m - y_m$.
$x = \dfrac{\text{Transmitted frequency}}{\text{Critical frequency}}$.
i, i_o = Angle of incidence of ray on layer.
α = Angle of elevation of ray at aerial.
c = Free-space velocity of electromagnetic waves.
P_T = Total power radiated by transmitter.
P_R = Available power from receiving aerial.
G_T, G_R = Power gain of transmitting and receiving aerials relative to an isotropic radiator.
λ = Wavelength.
r = Range of radar target.
A = Echoing area of radar target.
A_o = Echoing area of unit surface area of ground.
δt = Duration of transmitted pulse.
θ = Co-ordinate of azimuth at aerial.
θ_B = Azimuthal beamwidth of aerial.
F_o = Power-flux density at unit distance from transmitting aerial.
F_t = Power-flux density incident on ground.
R = Radius of earth.
a = Radius of spherical scattering object.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

(1) INTRODUCTION

There has been considerable interest in recent years in the weak, scattered echoes which can be received near a short-wave pulse transmitter operating above the vertical-incidence penetration frequency of the ionosphere. Eckersley[1] has studied the various modes of propagation of these echoes, and has distinguished two main categories. The first—echoes received at ranges from 80 to 500 km—he attributed to energy scattered from irregularities at E-region heights illuminated by direct radiation from the transmitter. These echoes, reported by Appleton,[2] have recently been extensively studied and the predominant sources have now been identified as the ionized trails left by meteors. This category will not be discussed in the paper. Echoes of the other type distinguished by Eckersley are received at ranges up to several thousand kilometres. He showed that they were returned from sources illuminated by rays reflected back towards the earth by the ionosphere, and concluded that the sources of these long-range echoes were also in the E region.

Edwards and Jansky[3] suggested as more probable that the sources were irregularities on the earth's surface, and this theory has since been supported by experimental evidence given by Kōno,[4] Dieminger,[5] Peterson[6] and Abel and Edwards.[7] Further evidence in favour of this theory is given in the present paper. Since 1945, theoretical and experimental investigations of long-range back-scatter have been described by several workers.[8,9,10,11]

The existence of long-distance scatter has suggested a new technique for investigating the ionosphere, since it should be possible to measure the skip distance in a chosen direction for any frequency by timing the delay of echoes arriving from that direction. Ionospheric sounding stations, which with established technique can measure only the characteristics of the ionosphere directly overhead, should by this new method be able to provide information about the layers above an area several thousand kilometres in diameter.

In a refinement of this technique developed by Villard and Peterson,[12] and independently by de Bettencourt,[13] a continuously rotating aerial is used in conjunction with a rotating time-base indicator. This enables the observer to see the areas from which echoes are received in map form on a plan-position display. Earlier experiments with a rotating aerial in conjunction with a Cartesian co-ordinate display had been reported by Kōno.[4]

If it is desired to measure the maximum usable frequency (m.u.f.) for one-hop propagation over a given distance, the frequency of transmission must be varied. The m.u.f. is then the frequency for which the measured skip distance equals the required distance. Since it is often possible also to detect scattered echoes returned from subsequent ground-reflection points along the path studied, the lengths of the individual hops in multi-hop propagation can be measured. Such measurements, which are essential if long-distance propagation is to be

understood, are very difficult to obtain by other methods, all of which require the co-operation of distant observers.

The work described in the paper began in 1949, when an investigation of long-distance back-scatter was begun at the Radio Research Station. The objects of the investigation were to find out more about the nature of scattered echoes and their sources and modes of propagation, and to investigate experimentally the accuracy and limitations of a back-scatter technique for skip-distance measurement.

In the paper, a theory is developed of long-distance scatter propagation, in which the relation between transmitted frequency, time delay and ground range of scattered echoes is found for a simplified model of the ionosphere. The model taken is that of a curved earth and a curved layer with parabolic ionic-density distribution, as discussed by Appleton and Beynon.[14] In their analysis the effect of the earth's magnetic field on the wave as it passes through the ionosphere is neglected. The same model is used to study the factors affecting the intensity of scattered echoes, including the important effect of ray focusing discussed by Peterson.[9]

In order to estimate the effect of the earth's magnetic field on scatter propagation, a study is made of the results of oblique-incidence pulse observations by other workers. So far, suitable experiments have been made only over short distances. However, the results make it possible to determine the effect of the field for short-range propagation and thus to remove any ambiguity when determining whether the scatter sources are on the ground or in the E layer.

(2) PLANE EARTH AND PLANE IONOSPHERE

The propagation of long-distance back-scatter will first be discussed for the simplified model of a plane earth and plane, thick ionosphere, with no terrestrial magnetic field present. Plane-earth theory should provide a satisfactory approximate solution to the true curved-earth problem for ground ranges up to 900 km if corrections are made for the effect of the earth's magnetic field. This simplified model has been studied by others,[5,6] but the theory is summarized here in order to make the subsequent work clear.

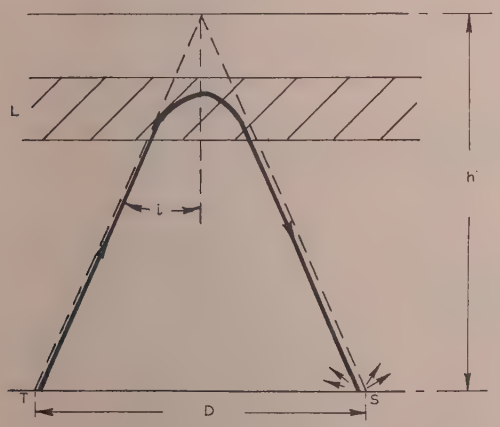


Fig. 1.—Ray path for ground scatter reflected by a plane layer.

Fig. 1 shows a typical ray path for a scattering source on the ground. One ray is shown which leaves the transmitter T and is bent by the ionized layer L so that it returns to the ground at S at a range D. If p' is the equivalent free-space path for the frequency f at an angle of incidence i, and h' is the virtual height of reflection for a frequency f_v at vertical incidence, such that

$$f_v = f \cos i \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

then, by Martyn's Theorem,

$$p' = \frac{2h'}{\cos i} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and, by Breit and Tuve's Theorem,

$$D = p' \sin i$$

Thus, by combining the last two equations,

$$D = 2h' \tan i \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In a pulse experiment the time delay of echoes from the scattering sources which are illuminated by these oblique rays can be measured. If the transmitter and receiver are close together, and if it is assumed that the outward and return paths are identical, the time delay is

$$t = \frac{2p'}{c}$$

where c is the velocity of light.

The time-base of the display may therefore be calibrated in kilometres and the length of the equivalent path read directly from the face of the cathode-ray tube. At any chosen frequency, it should be possible to receive scattered echoes from the ground at all angles of incidence for which rays are reflected by the layer, provided that the field strength is sufficient.

The pattern expected on the receiver time-base would consist of a mass of echoes having path lengths varying from a minimum, which would depend upon the height of reflection and the angle at which the layer is penetrated, up to a maximum which would be governed by the minimum field-strength detectable. The minimum path length is the most easily measured property of the scatter, so that it is of interest to calculate it from the above equations.

From eqns. (1) and (2), the general expression for the path length, from which the minimum path-length p'_{min} must be determined, is

$$p' = \frac{2h'}{f_v} \times f \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

provided that $f_v \leqslant f$ (since $\cos i \leqslant 1$).

This equation is most easily interpreted graphically, since it depends on the relation between h' and f_v, a physical property of the layer found by vertical-incidence sounding measurements. A typical curve of $2h'$ against f_v for a simple layer without taking into account the effect of the earth's magnetic field is shown in Fig. 2. The factor $2h'/f_v$ in eqn. (4) is given by the slope of a straight line from the origin to the point, or points, $(2h', f_v)$. The minimum value of $2h'/f_v$ is given by the tangent to the curve.

From eqn. (4) it follows, as Dieminger[5] has shown, that the relation between the minimum path-length of scatter and the transmitted frequency is given to the same scales by the tangent line, provided that $f \geqslant f_{tan\,i}$. For lower frequencies the relation is given by the $2h'/f_v$ curve itself. Fig. 2 shows the possible path lengths of scatter superimposed on the $2h'/f$ curve.

The path length for E scatter is found by subtracting the oblique distance from E layer to ground ($h_E/\cos i$) from the ground-scatter equivalent path, thus

$$p' = \frac{2h'}{\cos i} - \frac{h_E}{\cos i}$$

$$= \left(\frac{2h' - h_E}{f_v}\right) \times f \quad f_v \leqslant f \quad . \quad . \quad . \quad (5)$$

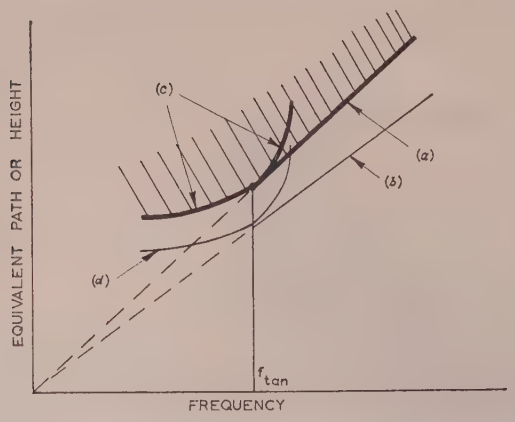


Fig. 2.—Calculated path-length for ground and E-layer scatter.
Possible path-lengths for ground scatter are represented by shaded areas.
(*a*) Minimum path-length for ground scatter.
(*b*) Minimum path-length for E scatter.
(*c*) $2h'/f_v$ curve from vertical-incidence data.
(*d*) $(2h' - h_E')/f_v$ curve.

The minimum path-length is again proportional to frequency for $f_v \leqslant f$, and is found by drawing the tangent from the origin to a curve of $(2h' - h_E)$ against f_v (Fig. 2).

The difference between the minimum-path-length lines for ground and E scatter provides a method for ascertaining whether scatter comes from the E region or from the ground. By means of a variable-frequency ionospheric recorder Dieminger obtained photographic records which showed scatter path-lengths following the law expected for ground scatter after allowance had been made for the effect of the earth's magnetic field.

Peterson[6] has carried out similar swept-frequency investigations and has presented the results of two such experiments. These show the observed path-lengths following the tangent law for ground scatter derived above. There is, however, an important difference between these results and Dieminger's, since Peterson obtained agreement with the law expected for the ordinary ray, which behaves approximately as if no magnetic field were present, whereas Dieminger found agreement with the law corrected to allow for the shorter path-length of the extraordinary ray. Since the difference between the corrected and uncorrected path lengths is comparable with the difference between the path lengths of ground and E scatter, it is a matter of importance to find which of these methods is accurate in order to establish the location of the scattering sources without ambiguity.

(3) EFFECT OF THE EARTH'S MAGNETIC FIELD ON THE EQUIVALENT PATH OF SCATTERED ECHOES

In Section 2 a simplified model of a plane earth and plane ionosphere with no terrestrial magnetic field present was examined, and a linear relation found between the minimum path-length of scatter and the transmitted frequency. It was assumed, however, that Martyn's theorem applied and, since it does not apply in the presence of the earth's magnetic field, a further examination is necessary before the theory can be applied to the practical problem.

The effect of the earth's magnetic field on a wave passing through an ionospheric layer is to split the wave into ordinary and extraordinary components and to distort the path of these components in a manner depending on the orientation of the incident wave and of the field vector. The calculation of such paths is extremely complex and is not justified for the purpose in view. The only practicable course for this work is to make an empirical allowance for the effect of the earth's magnetic field based on experimental pulse observations at oblique incidence.

Table 1 summarizes the available information on the accuracy of no-field theory when applied to the ordinary ray at oblique incidence, and also gives the observed separations between ordinary- and extraordinary-ray m.u.f.'s.

It will be seen that the agreement between calculated and observed ordinary-ray m.u.f.'s is very good, and that for distances up to about 800 km in a north-south direction the separation between ordinary- and extraordinary-ray m.u.f.'s is nearly the same as at vertical incidence. In the single record available for the 1 150 km east-west path the m.u.f. separation is about half the critical-frequency separation, but this must be

Table 1

OBLIQUE-INCIDENCE PULSE EXPERIMENTS

Authors and type of experiment	Distance from transmitter to receiver	Theory used to calculate m.u.f. from critical frequency	Difference between observed and calculated ordinary-ray m.u.f. Mean values	Ordinary-ray m.u.f. factor	Separation between ordinary-ray and extraordinary-ray m.u.f.	
					Oblique incidence	Vertical incidence
Farmer, Childs and Cowie.[15] Manual; $p'f$	km 464 Edinburgh–Cambridge (approximately N.–S.)	Plane ionosphere; Curved earth	% 0·47 Probable error ±0·13% (21 observations)	About 1·05	Mc/s 0·62 ± 0·013	Mc/s Nearly the same as oblique incidence
Beynon.[16] Manual and automatic; $p'f$	715 Burghead–Slough (approximately N.–S.)	Curved earth; Curved ionosphere; Parabolic layer	< 1 (110 observations)	1·25–1·42 through the year	0·66 ± 0·02	0·66 (Slough) 0·64 (Burghead)
Strohfeldt, McNicol and Gipps.[17] $p't$	763 Brisbane–Camden, N.S.W. (approximately N.–S.)	Curved earth; Curved ionosphere; Parabolic layer			0·65 ± 0·02	$\frac{1}{2}f_H = 0·66$
Sulzer and Ferguson.[18] (Single published record only)	1 150 Sterling, Va.–St. Louis, Mo. (approximately E.–W.)	CRPL transmission curve. Curved earth; curved ionosphere	1 (1 observation)	1·7	0·45 (1 observation)	0·82 at mid-point of path. (1 observation)

regarded as insufficient evidence, since unless several observations are made the effect of horizontal gradients may be unknown.

These results may be expressed in a different way by saying that, for ground ranges up to 800 km, the ordinary-ray skip distance agrees with that calculated from no-field theory, while for the extraordinary ray a given skip distance is obtained at a frequency $\frac{1}{2}f_H$ higher than that for the ordinary ray, f_H being the gyro frequency.

In considering the effect of the earth's magnetic field on observations on scatter, it is the equivalent path-lengths rather than the skip distances which are of interest. In the Burghead–Slough experiments Beynon obtained good agreement between the path length of the ordinary ray at oblique incidence and that calculated from no-field theory. Inspection of vertical-incidence $h'f$ records and the oblique-incidence records obtained by Beynon shows that a similar approximate empirical relation exists between the path lengths of the extraordinary and ordinary rays as between the skip distances, a given path-length being obtained for the extraordinary ray at a frequency $\frac{1}{2}f_H$ higher than for the ordinary ray.

Thus for a particular ionosphere and a transmitted frequency f, if

$$p' \text{ (ordinary)} = \phi(f, D)$$

then

$$p' \text{ (extraordinary)} = \phi(f - \tfrac{1}{2}f_H, D)$$

To obtain the minimum path-length for scatter at a particular frequency, f is kept constant in the above equations and D is varied by varying the angle of elevation of the transmitted ray until p' passes through a minimum. The minima of both expressions are the same, and their value is obtained for the same value of the frequency parameter. For the ordinary ray, from Section 2,

$$p'_{min} \text{ (ordinary)} = \frac{2h'_{tan}}{f_{tan}} \times f$$

so that for the extraordinary ray

$$p'_{min} \text{ (extraordinary)} = \frac{2h'_{tan}}{f_{tan}} \times (f - \tfrac{1}{2}f_H)$$

h'_{tan} and f_{tan} being found from the ordinary-ray vertical-incidence curve. If the ordinary- and extraordinary-ray $h'f$ curves are identical in shape, this is the equation of the tangent to the extraordinary-ray $2h'f$ curve, not from the origin, but from the point $\frac{1}{2}f_H$ on the frequency axis. This form of correction is the same as that used by Dieminger, so that the above argument confirms his modification of the tangent law for extraordinary-ray propagation. Fig. 3 shows the extraordinary-ray minimum-path-length curve found by this method for a parabolic layer with ionospheric parameters typical of the F2 region for a winter day at Slough.

In order to find the path length for E-region long scatter, $h_E/\cos i$ must be subtracted from the ground-scatter path-length as was done in Section 2. This necessitates a knowledge of the way in which the angle of elevation of the extraordinary ray varies with the path length.

Measurements of the angle of arrival of ordinary and extraordinary rays over a 700 km path have recently been made by Bramley[19] using the technique described by Bramley and Ross.[20] A study has been made of the results of experiments carried out when the ionization in the layer was falling steadily and the receiver was passing into the skip zone. This shows that the error introduced for these conditions by taking the angle of arrival of the extraordinary ray at a frequency f to be the same as that of the ordinary ray at $f - \frac{1}{2}f_H$ will not be more than 2° or 3° in about 40°. The slight uncertainty is caused by the random tilts of the layer which occur even on an ionospherically quiet day, but will introduce a negligible error into the path-length calculation for the short ranges considered.

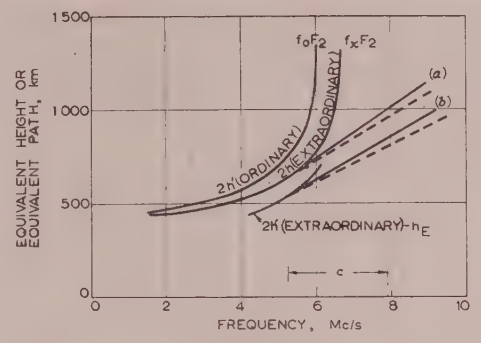


Fig. 3.—Minimum path-length of scatter for extraordinary-ray propagation.

(*a*) Ground scatter.
(*b*) E scatter.
(*c*) Range for which no-field theory is a good approximation.
———— Corrected for magnetic field.
– – – – – No-field theory.
Parabolic layer. $f_0F_2 = 6{\cdot}0$ Mc/s. $f_xF_2 = 6{\cdot}66$ Mc/s. $h_m = 350$ km. $y_m/h_0 = 0{\cdot}6$.

Thus the extraordinary-ray path-length for E scatter is

$$p'_E = p'_{ground} - \frac{h_E}{\cos i} = \frac{2h'}{f_v} \times (f - \tfrac{1}{2}f_H) - \frac{h_E}{f_v} \times (f - \tfrac{1}{2}f_H)$$

$$= \frac{2h' - h_E}{f_v} \times (f - \tfrac{1}{2}f_H)$$

h' and f_v again referring to the ordinary-ray $h'f$ curves.

The relation between minimum path-length and frequency is again linear, and is found by drawing a tangent from the point $\frac{1}{2}f_H$ on the frequency axis to the extraordinary-ray $2h'$ curve with the ordinates diminished by h_E (see Fig. 3). It will be seen from the figure that the error in applying the no-field theory to the extraordinary ray is very small for both ground and E scatter at the frequency corresponding to the tangent point, and rises to 4% at a frequency 50% above this. The path length of E scatter is 15% shorter than that of ground scatter, so that for this frequency range the error in calculating path length by using the simple theory will be small compared to the difference in path lengths of ground and E scatter.

To summarize this analysis of oblique-incidence experiments, it appears that no ambiguity will be introduced in the location of the sources of long-range scatter if the results are compared with a simplified theoretical relation. This relation is obtained by applying no-field theory to the extraordinary-ray vertical-incidence curve, and is valid within a restricted frequency range.

(4) LOCATION OF THE SCATTERING SOURCES RESPONSIBLE FOR LONG SCATTER

The analysis of the previous Sections provides a method of establishing, by swept-frequency observations, the location of the sources of long-distance back-scatter. In this Section the results of some swept-frequency ($p'f$) experiments are given, and further evidence relevant to the location of the scattering sources is discussed.

(4.1) Swept-Frequency Scatter Observations

In 1949 experiments were begun at the Radio Research Station with the object of establishing the location of the scatter sources. A brief account of this work has appeared elsewhere.[21]

In these experiments, pulse transmissions were made on a number of frequencies and the time delay of the first long-scatter echoes to return was measured for frequencies both below and above the critical frequency of the F2 region. The relation between the time delay of the scattered echoes and that of the vertical-incidence echoes obtained below the critical frequency was compared with the simple tangent law derived in Section 2.

The equipment consisted of a variable-frequency pulse-modulated transmitter with an automatically synchronized receiver, and covered the frequencies 1–17 Mc/s in one range of a manual or motor-driven tuning control. The power output of the transmitter was 10 kW with a pulse length of 140 microsec, and both the receiver and transmitter were coupled to one horizontal wide-band aerial by an aperiodic transmit-receive switch. The aerial had a substantially circular azimuthal polar diagram for high angles of elevation.

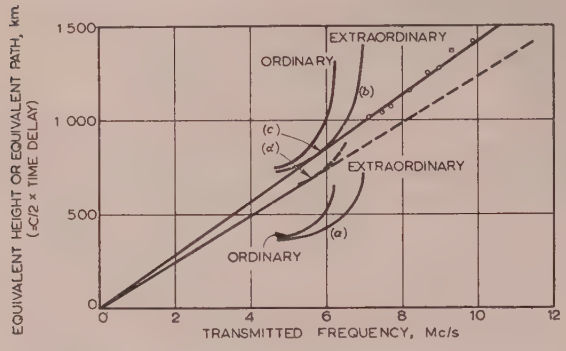


Fig. 4.—Predicted and observed path-lengths of leading edge of scatter group.

(*a*) 1F vertical-incidence h'/f curve.
(*b*) 2F h'/f curve.
(*c*) Tangent to 2 h'/f curve.
(*d*) Tangent to 2 $h' - h_E/f$ curve.
———— From ground (calculated).
– – – – – From E layer (calculated).
o Observed.
30th June, 1949, 0400 G.M.T.

Fig. 4 shows the results of one experiment plotted as a graph of equivalent path-length or height against frequency. It will be seen that the path lengths of the leading edge of the scatter pattern lie close to the tangent to the extraordinary-ray $2h'f$ curve (the theoretical law for ground scatter) and not to the line predicted for E-region scatter.

It has been assumed here that the tangent law applies to extraordinary-ray propagation up to 60% above the "tangent frequency." This assumption has been justified for approximately this range in the discussion of Section 3.

Although in this single experiment the observations agree with the theoretical law for ground scattering, this cannot be regarded as sufficient evidence, since the theory assumes that the layer is uniform over a large area surrounding the transmitter. It is therefore desirable to analyse the results of a number of experiments in order to eliminate the effect of random variations in the layer.

For ground ranges up to 900 km, where plane-earth theory is a good approximation, the observed minimum path-length of scatter and the corresponding frequency may be normalized with respect to the height and frequency co-ordinates of the tangent point. The normalized path-length and frequency for a number of $p'f$ experiments can then be plotted on a common graph of p'/h_{tan} against f/f_{tan}. This has been done for seven manual $p'f$ experiments made during the night of 29th–30th June, 1949, and twenty-two automatic records taken on 14th May, 1952, all at Slough.

Fig. 5(*a*) is a histogram obtained from the resulting dot diagram for the frequency range up to $1{\cdot}5f_{tan}$. Path lengths

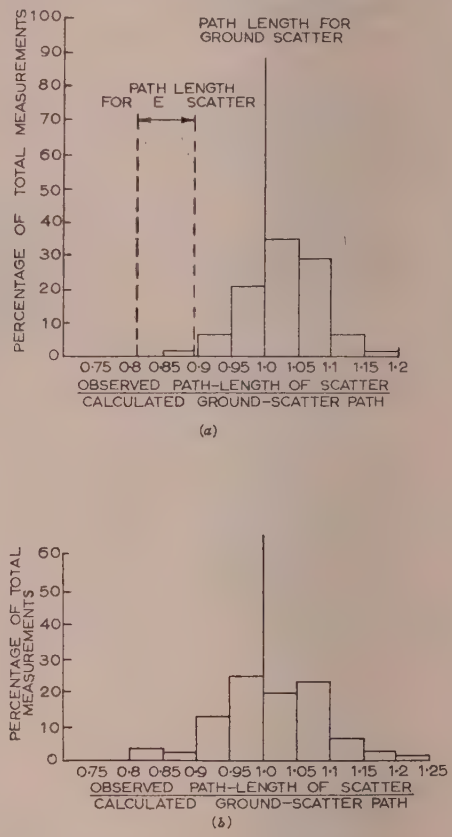


Fig. 5.—Histograms of observations to locate sources of scatter.

(*a*) Frequency $1{\cdot}0f_{tan}$–$1{\cdot}5f_{tan}$.
(*b*) Frequency $1{\cdot}5f_{tan}$–$2{\cdot}0f_{tan}$.

were observed both longer and shorter than the theoretical ground-scatter value with the mean at approximately +2%, whereas the path length for E scatter should have centred at approximately −15%. It will be noticed that a range of values has been put for the E-scatter path-length. This is because the path length for this mode is not a fixed percentage shorter than that for ground scatter but varies between the extremes of −11% and −18% for the different heights and critical frequencies observed in the various experiments.

The random differences between the calculated and observed path-lengths were to be expected, since the calculations assume that the ionosphere was the same at the reflection point for oblique scatter as at a point directly over the transmitter. This would require an ionosphere uniform over an area of 500 km radius—an improbable condition.

The histogram for the frequency range $1{\cdot}5f_{tan}$–$2{\cdot}0f_{tan}$ is given in Fig. 5(*b*), showing a poorer correlation than for $1{\cdot}0f_{tan}$–$1{\cdot}5f_{tan}$. The correlation would be expected to decrease at longer ranges, since the reflection points were further away and the random differences in the ionosphere consequently likely to be greater. Errors caused by the earth's magnetic field, by earth curvature and by ionosphere gradients should also have been greater for the longer distances.

It is concluded from these experiments that the sources of the scattered echoes were on the ground and not in the E region. The conclusions of other investigators are therefore confirmed.

(4.2) Further Evidence for Ground Scatter

It was mentioned in Section 1 that, in addition to the long scatter regularly observed with a high-power transmitter, transient echoes are seen at short ranges (100–500 km). These echoes have been identified by several workers as being caused by the ionized trails left by meteors passing through the E region. The variation of amplitude with time of these echoes is characteristic, a pulse-shaped echo appearing suddenly and dying away with a cyclic flutter of amplitude in a period of one or more seconds.

In the course of back-scatter observations, transient, fluttering echoes of the same type, but generally of smaller amplitude, are occasionally seen at much longer ranges, the echoes appearing 200–300 km in front of the main scatter group on the receiver time-base. The variation of amplitude with time is very different from that of the main scatter group, which consists of a mass of peaks slowly fading at random. It is concluded that the transient echoes are due to meteors in the E region, and the fact that they are seen some distance in front of the main scatter group is a further confirmation that the main scatter comes from the ground and not from the E region.

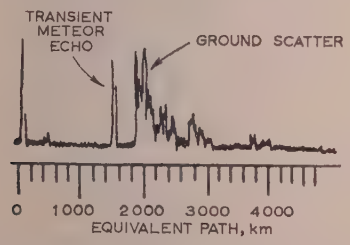


Fig. 6.—Example of long-range meteor echo.

Fig. 6 is reproduced from a photograph of the time-base display taken at a time when a long-range meteor echo of this type was present. A rotating aerial was used, directed towards the south, where the shortest range of scatter was observed. By this means the possibility that the meteor echo was very intense and came from a direction away from the aerial beam, where the skip distance was shorter, is precluded.

So far, consideration has been confined to ground or E scatter reflected by the F layer. However, echoes are frequently seen, particularly in summer, which are due to energy scattered by the ground and reflected on the outward and return journey by the Es layer. The justification for assuming that the echoes are from the ground is given by photographic $p'f$ records showing the scattered echo emerging from the 2Es vertical-incidence echo, and increasing with range in accordance with the tangent relation found in Section 1. A record showing this phenomenon has been given by Dieminger,[5] and similar records have been obtained at the Radio Research Station. With echoes of this type the minimum time delay does not agree so consistently with the tangent law, probably because of the very irregular nature of the Es clouds; these are often small in extent and therefore do not have a constant critical frequency over the area of several hundred kilometres necessary for the analysis to be applicable.

Kōno[4] has measured the angle of arrival of this type of scatter and compared the measured angles with the corresponding path length. The relation was in close agreement with the expected law for ground scatter reflected by a layer at a height of 100 km.

(5) PLANE EARTH AND PLANE LAYER WITH PARABOLIC IONIC-DENSITY DISTRIBUTION

Since the various experimental results which have been discussed all indicate that long-distance back-scatter is predominantly from the ground, only this type of scatter propagation will be considered in the theoretical analysis following. No-field theory only will be considered.

The equations given in Section 1 make it possible to calculate the ground range of scattered echoes from the observed equivalent path if the relation between frequency and height of reflection at vertical incidence is known. It is useful to specify a model ionosphere having characteristics similar to those observed, and to calculate the relation between the scatter parameters for this model. The model taken here is the parabolic layer considered by various authors. In the nomenclature of Appleton and Beynon[14] for such a layer over a plane earth, the virtual height of reflection at vertical incidence is

$$h'_o = \frac{1}{2} y_m x \log_\epsilon \frac{1+x}{1-x} + h_0$$

where
$h_0 = h_m - y_m$
h_m = Height of maximum ionization
y_m = Semi-thickness
$x = \dfrac{\text{Transmitted frequency}}{\text{Critical frequency}}$

At oblique incidence

$$p' = y_m x \log_\epsilon \frac{1 + x\cos i_o}{1 - x\cos i_o} + \frac{2h_o}{\cos i_o}$$

$$D = p' \sin i_o$$

where i_o = angle of incidence on "bottom" of layer (height h_o).

It is convenient to normalize the distances with respect to the height of maximum ionization, h_m. The above equations then become, after simplification,

$$\frac{p'}{h_m} = 2\frac{xy_m/h_o}{1 + y_m/h_o} \times \text{arc tanh}\,(x\cos i_o) + \frac{2}{\left(1 + \dfrac{y_m}{h_o}\right)\cos i_o} \quad . \quad (6)$$

$$\frac{D}{h_m} = \frac{p'}{h_m} \sin i_o \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Fig. 7 shows the relation between the normalized path-length and ground range for various values of the frequency parameter x. Scales of p' and D in kilometres have been added for one particular value of h_m, namely 350 km, to show the orders of distance to be expected. The curves have been calculated for a value of y_m/h_o, the parameter defining layer thickness, equal to 0·4.

Because of the simple relation, eqn. (7), between D, p' and the angle of incidence, straight lines radiating from the origin represent fixed angles of incidence. For a given path-length, therefore, the ground range and angle of incidence may be read direct from the chart, if the value of x is known.

The effect of varying the layer thickness for a given value of h_m is shown in Fig. 8.

It will be seen from these curves that at short distances there is a considerable difference between the skip distance (minimum ground range) and the ground range corresponding to minimum path-length. Physically, this means that the leading edge of the scatter pattern seen in a pulse experiment comes from beyond the skip distance and is not reflected from the same part of the ionosphere as the skip ray. However, at ranges greater than 500 km these differences become small and comparable with the range resolution of the pulse equipment. The curves have not

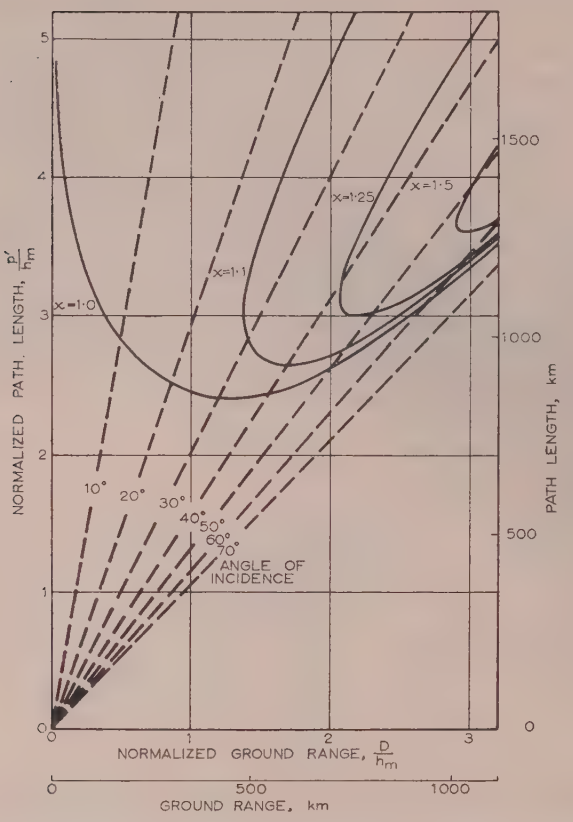


Fig. 7.—Relation between ground range and path length for plane earth.

Parabolic layer. $y_m/h_o = 0{\cdot}4$

$$x = \frac{\text{frequency}}{\text{critical frequency}}$$

Ranges and path-length scales in kilometres are for $h_m = 350$ km only.

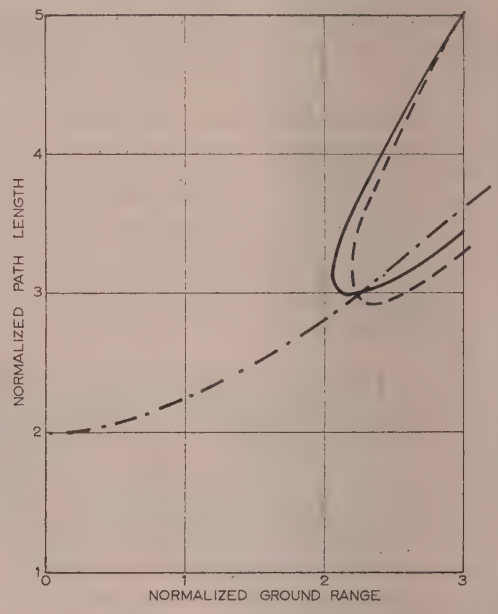


Fig. 8.—Effect of layer thickness on relation between ground range and path length.

$x = 1{\cdot}25$.
$y_m/h_o = 0$.
$y_m/h_o = 0{\cdot}4$.
$y_m/h_o = 0{\cdot}8$.

been extended to large distances as flat-earth theory becomes inadequate for ground ranges greater than about 900 km.

Fig. 9 shows the minimum path-length and skip distance plotted against frequency, showing the linear path-length/frequency relation, and also showing the very rapid increase of skip distance with frequency just above the critical frequency. This latter property means that considerable errors are likely to occur if an attempt is made to deduce the skip distance from the equivalent path of the leading edge of the scatter pattern for frequencies less than 10% above the critical frequency. With $h_m = 350$ km, for example, an increase of 60 km in equivalent path can correspond to a change of skip distance from 0 to 370 km. However, for frequencies more than 20% above the critical frequency, corresponding to skip distances of more than 500 km, it should be possible to deduce the skip distance from the minimum path-length with good accuracy.

(6) CURVED EARTH AND CURVED, PARABOLIC LAYER

The curvature of the earth and ionosphere makes the calculation of skip distance and equivalent path considerably more complicated. The normalization process used in the last Section is no longer possible and separate curves must be calculated for each layer height. Expressions for the ground range, equivalent path and angle of incidence, i_o, on the layer are given in Section 14.1. The angle of elevation, α, of the ray

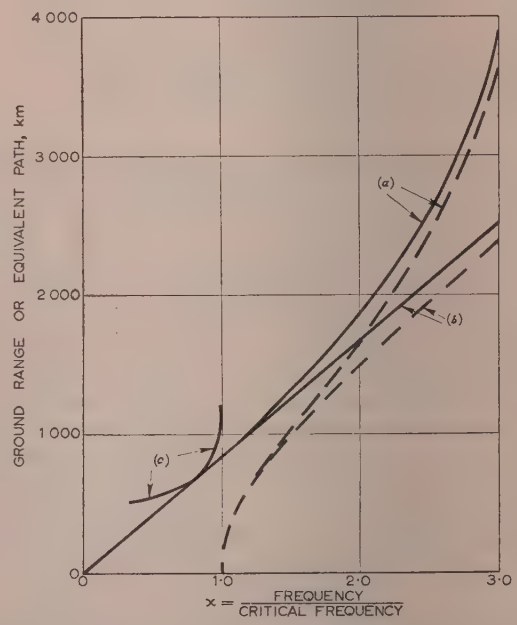


Fig. 9.—Variation of skip distance and minimum path-length with frequency for plane and curved earths.

(*a*) Curved earth.
(*b*) Plane earth.
(*c*) Vertical-incidence $2h'/f$ curve.
– – – – – Skip distance.
———— Minimum path-length.
Parabolic layer $\begin{cases} h_m = 350\,\text{km}. \\ y_m/h_o = 0{\cdot}4. \end{cases}$

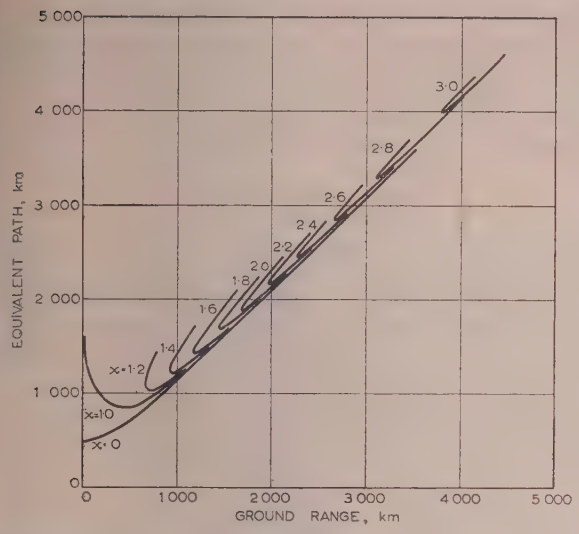


Fig. 10.—Relation between ground range and equivalent path for a curved earth and parabolic layer.

$h_m = 350$ km $y_m/h_o = 0{\cdot}4$.

$x = \dfrac{\text{transmitted frequency}}{\text{critical frequency}}$

at the ground is also given, since for a curved earth this is no longer the complement of i_o. The curves of p' and D in Fig. 10 have been calculated from these expressions for a layer having $h_m = 350$ km and $y_m/h_o = 0{\cdot}4$. The effect of varying the layer thickness for constant h_m is small, so that it is sufficient for interpreting scatter results to use a set of curves for constant y_m/h_o and various values of h_m.

The curves are of the same form as those for a plane earth, except that both ground range and path length increase more rapidly with frequency as the limiting condition ($x = 3{\cdot}16$) for the escape through the layer of the ray with zero angle of elevation is approached. This is shown more clearly in Fig. 9, where the minimum path-length and skip distance are plotted against frequency together with the corresponding curves for a plane earth.

These curves give the information needed to determine skip distance from observations of back-scatter echoes, but a simpler presentation is desirable. The quantity which can be observed with a pulse sounder is the time delay of the echo forming the leading edge of the scatter pattern, this delay being measured for convenience as an equivalent free-space range. This range is the minimum equivalent path, p'_{min}, given in the above curves, and the relation required to determine skip distance is therefore that between p'_{min} and the skip distance D_{skip}. Since these quantities are nearly equal at long range, it is more convenient to plot against p'_{min} the correction to be subtracted from it to obtain skip distance. Curves of this correction against p'_{min} are given in Fig. 11 for a range of layer parameters corresponding to normal F-layer conditions ($h_m = 350$, 310 and 270 km, and $y_m/h_o = 0{\cdot}4$). It will be seen that at long ranges D_{skip} can be found from p'_{min} to very good accuracy (error $< \pm 3\%$), without knowledge of the true value of h_m, by assuming a mean value. It is interesting to note that for p'_{min} greater than 1 700 km an error of less than 1% is introduced by using a simple linear relation of the form $D_{skip} = p'_{min} - K$, where K is a constant for any particular value of h_m.

The scatter technique thus appears to provide a good method for finding skip distance at a particular frequency, especially at long range, and thus satisfies an important practical need. It is

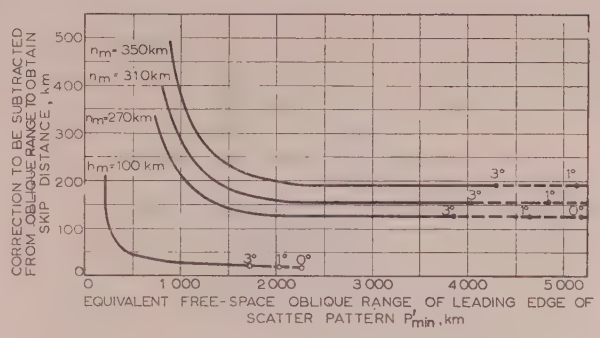


Fig. 11.—Chart for determination of skip distance from back-scatter range.

The figures given on the curves indicate the angle of elevation of the skip ray.
Layer parameters: F_2 layer. $y_m/h_o = 0{\cdot}4$, $h_m = 350$ km, 310 km, 270 km.
E_s layer. $y_m/h_o = 0$, $h_m = 100$ km.

not, however, a substitute for vertical-incidence sounding for physical investigation of the ionosphere, because scatter observations do not yield unique values for the height and critical frequency of a layer. For example, if p'_{min} is observed to be 1 890 km, and h_m is known to lie somewhere between 270 and 350 km, then D_{skip} is 1 630 km $\pm$ 3%. For a transmitted frequency of 15 Mc/s, the layer characteristics may be defined by ($f_o = 6$ Mc/s, $h_m = 270$ km) or ($f_o = 7{\cdot}5$ Mc/s, $h_m = 350$ km) or by corresponding pairs of values between these. The error in determining critical frequency is therefore $\pm 11\%$, which is considerably larger than the error in finding skip distance.

For the interpretation of ground scatter reflected by Es, a correction curve to obtain D_{skip} from p'_{min} for a thin layer ($y_m/h_o = 0$) at a height of 100 km has been included in Fig. 11. The limiting ground range for this mode is 2 240 km with an m.u.f. factor, x, of 5·7. M.U.F. factors should be used with caution for reflections by Es ionization because of the limited area and partially reflecting characteristics of the clouds.

To determine the limiting range for one-hop scatter with a particular aerial system it is necessary to know the angle of elevation, α, of the ray at the aerial. Accordingly the path lengths for $\alpha = 3°$ and $\alpha = 1°$ are shown in Fig. 11.

(7) AMPLITUDE OF SCATTERED ECHOES FROM THE RADAR EQUATION

In studies of the strength of long scatter made by other workers,[12] no marked relation has been reported between the strength of echoes and the type of terrain from which they originated, with the possible exception of echoes from mountainous regions.[5] Observations which have been made so far at the Radio Research Station confirm this lack of dependence on terrain. It appears that any such differences are masked by the large variations in echo amplitude introduced by the ionosphere. This is surprising, in view of the large differences between sea and land echoes noticed with airborne centimetric radar equipment, and of the large differences in the nature of the scattered signals associated with two-hop one-way propagation reported by Miya, Kobayashi and Wakai,[22] and by Bramley.[19]

For this reason it is interesting to consider the various factors affecting the strength of long scatter. The analysis is also useful when the accuracy of the scatter technique of skip-distance measurement is studied theoretically. The methods of calculating echo intensity can be used in this problem to find the shapes of the scatter patterns expected when the skip distance is changing across the aerial beam.

The most convenient approach to scatter problems is that of the radar equation.[23,24] According to this, the echo power available at the terminals of the receiving aerial of a radar system from a target at range r in free space is given by

$$P_R = \frac{P_T G_T}{4\pi r^2} \frac{A}{4\pi r^2} \frac{G_R \lambda^2}{4\pi}$$

where P_R = Available power from receiving aerial,
P_T = Total power radiated by the transmitter,
G_T = Power gain of transmitting aerial,
G_R = Power gain of receiving aerial (both gains relative to an isotropic radiator),
λ = Wavelength,
A = Echoing area of target.

This equation can be applied directly to the problem of a plane, thin ionosphere over a plane earth. Fig. 12 shows a pencil of

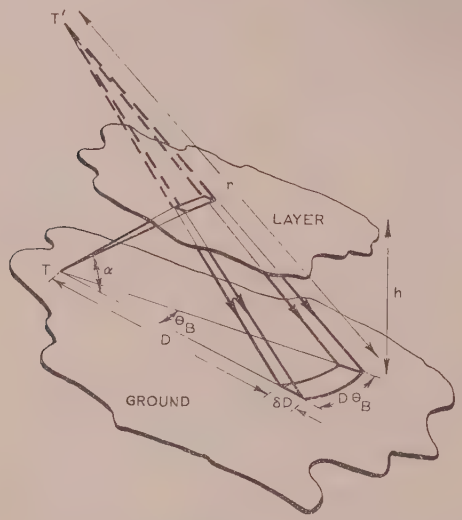


Fig. 12.—Ground area illuminated by a ray pencil reflected from a thin, plane layer.

rays leaving a transmitter T and being reflected back to the ground by such a layer. The field at the ground will be the same as that from an image transmitter T′ at a height $2h$ above T, where h is the height of the layer. The problem is therefore the same as that of ground echoes from an aircraft radar.

For a pulse of length δt, echoes will be received at any instant from an area on the ground for which the azimuthal angle subtended is equal to the beam width of the aerial and over which the oblique range changes by $\frac{1}{2}c\delta t$, where c is the free-space velocity of light. In the nomenclature of the figure, the ground area illuminated is

$$D\theta_B \delta D$$

For short pulses we can put

$$\delta D = \frac{\frac{1}{2}c\delta t}{\cos\alpha} = \frac{\delta r}{\cos\alpha}$$

To find the power scattered from this area, the total echoing area of the scattering sources upon it must be determined. A convenient model for a scattering source is the sphere, which may be represented for vertically polarized waves by a hemispherical boss upon perfectly conducting ground. The echoing area of a sphere is πa^2, provided that the radius, a, is large compared with a wavelength. Published curves[23,24] show that even if this criterion is not fulfilled the echoing area oscillates about a value of πa^2 as the radius is varied, the smallest radius to give this value being $a = 0{\cdot}1\lambda$.

Hence for n hemispherical bosses per unit ground area, the total echoing area of the illuminated zone is

$$A = n\pi a^2 D\theta_B \frac{\delta r}{\cos\alpha}$$

It is of interest to calculate from these equations the required density of scattering objects on the ground to account for the amplitude of echoes received by sporadic E reflection. It is to be expected that the most intense echoes will be received when a large sporadic E cloud is in the aerial beam and when the reflection coefficient of the cloud approaches unity.

An example of high-intensity echoes was observed at 1 100 G.M.T. on 13th July, 1953, using a rhombic aerial directed towards the west. The parameters in the above equations had the values

$P_T = 10^5$ watts $\quad P_R = 4{\cdot}4 \times 10^{-10}$ watt $\quad \lambda = 17$ m
$r = 7 \times 10^5$ m $\quad D = 6{\cdot}7 \times 10^5$ m $\quad \alpha = 16°$ $\quad \delta t = 120$ microsec

Measured aerial characteristics were not available, but the calculated values were

$G_T = G_R = 63$. $\quad \theta_B = 15°$ (0·262 rad) between half-power points.

Substitution of these values in the equations gives the area of the illuminated strip as $3{\cdot}29 \times 10^9$ m^2, the total power incident on it as 925 watts and the power reradiated (assumed isotropic) as 1·87 watts. If bosses of radius $0{\cdot}1\lambda (= 1{\cdot}7$ m) are considered, n is $6{\cdot}1 \times 10^{-5}$/m^2. A density of 61 bosses, each 1·7 m high, per square kilometre would therefore be sufficient to give the observed echo amplitude. Allowance must, however, be made in this calculation for the absorption of energy in the D region. Calculation from vertical-incidence absorption data gives an attenuation of 6·3 dB on both outward and return paths due to this effect. This means that the echoing area must be increased 18·2 times to return the same energy to the receiver, and leads to a density of about 1 000 bosses/km^2.

It should be stressed that this is necessarily an approximate calculation because no measured values are available for the aerial characteristics nor for the size of the Es cloud. However, accepting these limitations, the result suggests that the irregularities present on land and sea are sufficient to account for the strength of long-scatter echoes.

The simple analysis above is not applicable to propagation over greater distances, for which the arriving rays approach grazing incidence. For this condition vertically polarized waves are incident below the pseudo-Brewster angle and tend to undergo a phase reversal on reflection. The reflected rays reaching an object above ground level would then cancel instead of reinforcing the direct rays. The value of the echoing area of a hemispherical boss on imperfect ground would thus fall from πa^2, for waves incident above the pseudo-Brewster angle, to zero at grazing incidence, and would be expected to vary as $\sin^4\alpha$ for small angles of elevation.

(8) EFFECT OF FOCUSING AND ABSORPTION ON THE STRENGTH OF ECHOES

In the above discussion spatial attenuation only has been considered. In long-distance propagation by E- or F-layer reflection, however, an important factor is the focusing of rays caused by the curvature of the layer and by change of height of reflection with angle of incidence; this has been considered by various writers.[9,25,26] Expressions suitable for use with the

radar equation are derived in Section 14.2. The method adopted is to relate the energy flux, F_i, incident on the ground to the flux, F_o, at unit distance from the transmitter, making use of the property of a ray pencil that the energy flow through any cross-section is the same. The radar equation can then be adapted for propagation by a curved, thick ionosphere by putting it in the form

$$P_R = \frac{P_T G_T}{4\pi}\left(\frac{F_i}{F_o}\right)\frac{A}{4\pi}\left(\frac{F_i}{F_o}\right)\frac{G_R\lambda^2}{4\pi} \quad . \quad . \quad . \quad . \quad (8)$$

Allowance may be made for absorption of the waves in the D layer by multiplying the term F_i/F_o by the appropriate attenuation factor.

Before the results of a calculation of this kind for a parabolic layer and curved earth are examined, a few important focusing properties of a thick layer will be illustrated by means of the plane-earth model.

(9) RAY FOCUSING BY A PLANE, THICK IONOSPHERE

For a thin layer, where the height of reflection of a ray is independent of the angle of elevation, it has been shown that a virtual image of the transmitter is formed at a height $2h$ vertically above the transmitter itself. For a thick layer the same is true for rays of varying azimuth and constant angle of elevation, since they also are reflected at a constant height, but it is not true for rays in the same vertical plane. For these rays, as the angle of elevation of a ray pencil changes, so the point of focus moves along a caustic curve.

This property is illustrated in Figs. 13(a) to 13(d). For low-angle rays the image T′ lies close to the thin-layer position, but for higher-angle rays it moves down along the caustic curve formed by the envelope of the downcoming rays, becomes a real image as it passes through the layer, and moves down through the ground at the skip distance. The point of focus on the ground and on any other plane between it and the layer is thus at the inner edge of the illuminated zone on that plane.

By simple ray theory the energy flux at an image point is infinite, so that the field should be infinite along the whole caustic. This is an example of the breakdown of ray theory on a caustic mentioned by Bremmer.[26] The theory is, however, adequate at greater distances, and for calculation of the field at the ground it is sufficient to carry out the analysis by ray theory first, and then to consider the corrections necessary near the skip distance.

(10) ECHO AMPLITUDE FOR A CURVED EARTH AND CURVED, THICK IONOSPHERE

The energy flux incident on the ground for a curved earth and curved parabolic layer with $h_m = 350$ km and $y_m/h_o = 0{\cdot}4$ has been calculated from the expressions derived in Section 14.2. In Fig. 14 the incident energy is plotted against ground range for various values of the frequency parameter x. Only the flux due to the low-angle ray is shown, since the high-angle-ray energy falls off very rapidly beyond the skip distance.

The intensity at the skip distance has been calculated from the expressions given by Bremmer (see Section 14.2). Bremmer shows that in the skip zone a little downcoming energy reaches the ground by diffraction, but the intensity rises rapidly at the edge of the zone and reaches a maximum at the skip distance. Beyond this distance, fringes are formed by interference between high-angle and low-angle rays, but these decrease rapidly in magnitude as the high-angle energy becomes negligible compared with the low-angle energy.

The figure shows that the focusing of the energy in the neighbourhood of the skip distance should be considerable, the power flux in certain cases being ten times that for a thin layer.

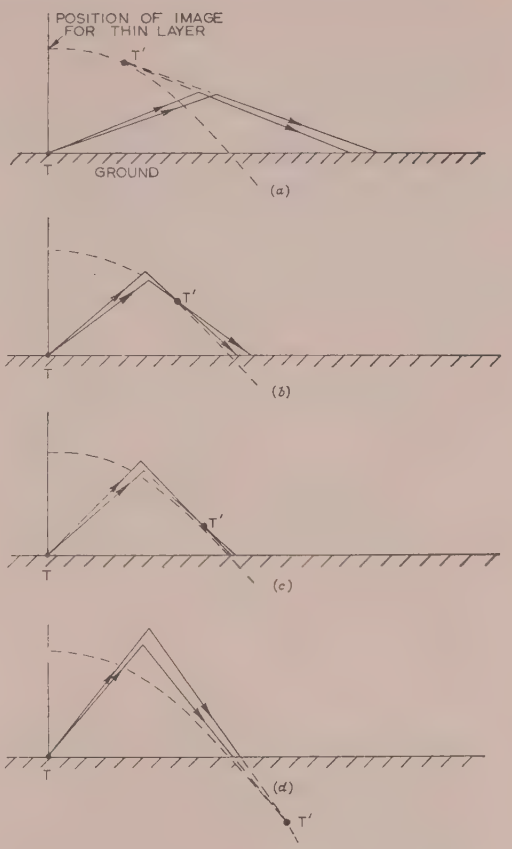


Fig. 13.—Focusing of rays by a thick layer.

Frequency constant; angle of elevation of ray pencil increasing from (a) to (d).
– – – – – Locus of T′, the image of transmitter T.
(Envelope of downcoming rays.)

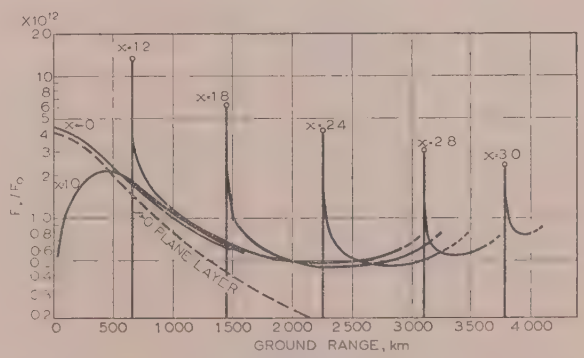


Fig. 14.—Variation of power flux incident on ground with distance from transmitter.

(F_o = Equivalent power flux at 1 metre from transmitter. For 1 kW and isotropic radiator, $F_o = 1\,000/4\pi$ watts/m².)
Curved earth, curved parabolic layer. $h_m = 350$ km. $y_m/h_o = 0{\cdot}4$.
For calculation of F_i/F_o at skip distance, a critical frequency of 6 Mc/s is assumed.
$x = \dfrac{\text{transmitted frequency}}{\text{critical frequency}}$

The amount of focusing caused by layer curvature can be seen from the figure by comparison of the incident energy for a curved layer with that for a plane layer (dotted curve). Intense focusing occurs for near-tangential rays, although the angle of elevation at which it takes place is so low that it will be counter-

acted by the poor low-angle radiation of practical aerial systems. For this reason the power-flux curves have not been continued to ranges for which the angle of elevation is less than 2°.

To calculate the amplitude of echoes from a pulse transmission it is necessary to calculate the total energy received at a particular instant. As in the thin-ionosphere problem, this is found by summing the energy returned from an area of ground for which the equivalent path changes by $\delta p' = \frac{1}{2}c\delta t$, corresponding to a change of ground range δD. For a thick ionosphere, in the neighbourhood of the minimum path-length condition, δD is considerably increased, resulting in an increase of illuminated area and an enhancement of echo strength. This effect, which Peterson[9] has called time-focusing, is taken into account in the following analysis.

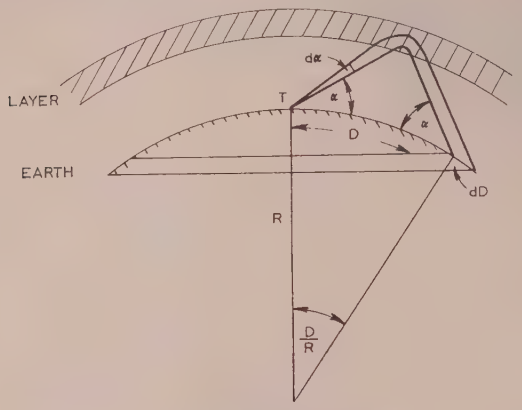


Fig. 15.—Ground area illuminated by a ray pencil for curved earth and curved, thick layer.

Considering an elemental strip on the ground (Fig. 15), the area subtended by an azimuthal angle $d\theta$ is

$$R \sin\left(\frac{D}{R}\right) d\theta dD$$

Let the echoing area of unit surface area of the ground be A_0. (For the model of hemispherical bosses $A_0 = n\pi a^2$.) The echoing area of the strip is then

$$dA = A_0 R \sin\left(\frac{D}{R}\right) d\theta dD$$

Owing to the rapid variation of the field in the immediate neighbourhood of the skip distance it is not justifiable, for pulse widths of 100 microsec or more, to assume that the energy flux is constant over the area of ground from which echoes are simultaneously received. The energy returned at any instant must be expressed as an integral over the area illuminated.

We must therefore put

$$P_R = \int \frac{P_T G_T}{4\pi}\left(\frac{F_i}{F_o}\right)\frac{dA}{4\pi}\left(\frac{F_i}{F_o}\right)\frac{G_R\lambda^2}{4\pi}$$

$$= \int_{\theta=0}^{\theta=2\pi}\int_{D-\delta D}^{D} \frac{P_T G_T}{4\pi}\left(\frac{F_i}{F_o}\right)^2 \frac{G_R\lambda^2}{4\pi}\frac{A_0}{4\pi} R \sin\left(\frac{D}{R}\right) d\theta dD$$

Assuming that the aerial gain does not change appreciably for the small change in angle of elevation corresponding to δD, and that the change in $\sin D/R$ is also small,

$$P_R = \frac{P_T\lambda^2 R}{64\pi^3}\int_{\theta=0}^{\theta=2\pi} d\theta\left[G_T G_R A_o \sin\left(\frac{D}{R}\right)\int_{D-\delta D}^{D}\left(\frac{F_i}{F_o}\right)^2 dD\right] \quad . \quad (9)$$

This expression can be used for the general problem of finding echo amplitude when the skip distance and the scattering properties of the terrain are varying across the aerial beam.

In the particular case where these can be assumed constant, and assuming that the same aerial is used for transmitting and receiving, the expression is

$$P_R = \frac{P_T G^2\lambda^2}{64\pi^3} R \sin\frac{D}{R} A_o\theta_B \int_{D-\delta D}^{D}\left(\frac{F_i}{F_o}\right) dD \quad . \quad . \quad (10)$$

where θ_B is the effective power beamwidth of the aerial, normally taken between half-power points.

Integration of these equations is most conveniently carried out graphically by measuring the area of a strip of width δD under the curve of $(F_i/F_o)^2$ against D, the value of δD necessary at any range being found from $\delta p'$ by reference to $p'D$ curves.

It should not be assumed that eqn. (10) is always a sufficient approximation. In practice a considerable change of skip distance across the aerial beam may occur. If it does, the energy from the neighbourhood of the skip distance will return at different times from different directions and will not add in the receiver. Since the greater part of the energy appears to come from the neighbourhood of the skip distance, a considerable drop in signal strength is to be expected if addition of the energy returning from different directions does not take place. The effect of this on the shape of the scatter pattern is discussed in a separate paper.[27]

(11) CONCLUSIONS

A theoretical study of the propagation of long-range back-scatter has shown that, if the scattering sources are on the ground, it should be possible to determine the skip distance by scatter observations alone to a reasonably high degree of accuracy, especially at long range (over 1 000 km). The scatter measurements do not, however, yield unique values of layer height and critical frequency.

The conclusion of various workers that long-range scatter originates at irregularities on the earth's surface has been confirmed by experiment. It has been demonstrated that no ambiguity is introduced into the result of this experiment by the effect of the earth's magnetic field.

Methods of calculating echo intensity for a thin reflecting layer have been given, and used to estimate the irregularities on the earth's surface required to give the observed echo intensity. The result, though necessarily approximate, suggests that the irregularities present on land and sea are fully sufficient to account for the echoes.

For a thick reflecting layer it is necessary to consider the effect of ray focusing on the echo intensity. The increase in the power flux incident on the ground due to focusing has been calculated from ray theory, except for the neighbourhood of the skip distance, for which a wave-theory method given by Bremmer has been used. The calculation shows that focusing can increase the power flux at the skip distance to ten times that obtained with a thin layer, corresponding to a threefold increase in field strength.

The complete analysis can be used to calculate the shape of the echo pattern on the receiver time-base when the aerial response and the skip distance are changing with azimuth. This is an important practical problem, since the accuracy of skip-distance measurement in a chosen direction depends on the extent to which echoes from unwanted directions can be rejected by the aerial.

The influence of the aerial polar diagram on the accuracy of skip-distance measurement has been considered in the analysis

of some further experimental results obtained with a high-power transmitter and various directional aerials. This work is described in a separate paper.[27]

(12) ACKNOWLEDGMENTS

The author wishes to acknowledge the help and advice of Mr. A. F. Wilkins, Mr. W. R. Piggott and Mr. L. T. J. Martin. Mr. C. G. McCue, of the Australian Department of Supply, co-operated, and assisted in the theoretical calculations.

The work described was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(13) REFERENCES

(1) Eckersley, T. L.: "Analysis of the Effect of Scattering in Radio Transmission," *Journal I.E.E.*, 1940, **86**, p. 548.

(2) Appleton, E. V., Naismith, R., and Ingram, L. J.: "British Radio Observations during the Second International Polar Year 1932–33," *Philosophical Transactions of the Royal Society*, A, 1937, **236**, p. 254.

(3) Edwards, C. F., and Jansky, K. G.: "Measurements of the Delay and Direction of Arrival of Echoes from Near-By Short-Wave Transmitters," *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 322.

(4) Kōno, T.: "Experimental Study on Scattered Echoes," Report of Ionospheric Research in Japan, 1950, **4**, No. 3.

(5) Dieminger, W.: "The Scattering of Radio Waves," *Proceedings of the Physical Society*, B, 1951, **64**, p. 142.

(6) Peterson, A. M.: "The Mechanism of F-Layer Propagated Back-Scatter Echoes," *Journal of Geophysical Research*, 1951, **56**, p. 221.

(7) Abel, W. G., and Edwards, L. C.: "The Source of Long-Distance Backscatter," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 1538.

(8) Benner, A. H.: "Predicting Maximum Usable Frequency from Long-Distance Scatter," *ibid.*, 1949, **37**, p. 44.

(9) Peterson, A. M.: "The Interpretation of Long Scatter Echo Patterns," Conference on Ionosphere Research, Pennsylvania State College, 1949.

(10) Hartsfield, W. L., Ostrow, S. M., and Silberstein, R.: "Back-Scatter Observations by the Central Radio Propagation Laboratory—August 1947 to March 1948," *Journal of Research of the National Bureau of Standards*, 1950, **44**, p. 199.

(11) Hartsfield, W. L., and Silberstein, R.: "A Comparison of C.W. Field Intensity and Backscatter Delay," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1700.

(12) Villard, O. G., Jr., and Peterson, A. M.: "Scatter-Sounding: A Technique for the Study of the Ionosphere at a Distance," Transactions of the Institute of Radio Engineers Professional Group on Antennas and Propagation, August, 1952, PGAP-3.

(13) de Bettencourt, J. T.: "Instantaneous Prediction of Ionospheric Transmission Conditions by the Communication Zone Indicator," *ibid.*

(14) Appleton, E. V., and Beynon, W. J. G.: "The Application of Ionospheric Data to Radio-Communication Problems: Part 1," *Proceedings of the Physical Society*, 1940, **52**, p. 518.

(15) Farmer, F. T., Childs, C. B., and Cowie, A.: "Critical Frequency Measurements of Wireless Waves reflected obliquely from the Ionosphere," *ibid.*, 1938, **50**, p. 767.

(16) Beynon, W. J. G.: "Propagation of Radio Waves," *Wireless Engineer*, 1948, **25**, p. 322.

(17) Strohfeldt, M., McNicol, R. W. E., and Gipps, G. de V.: "Ionospheric Measurements at Oblique Incidence over Eastern Australia," *Australian Journal of Scientific Research*, 1952, **5**, p. 464.

(18) Sulzer, P. G., and Ferguson, E. E.: "Sweep-Frequency Oblique-Incidence Ionosphere Measurements over a 1 150 km Path," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1124.

(19) Bramley, E. N.: "Some Comparative Directional Measurements on Short Radio Waves over Different Transmission Paths," *Proceedings I.E.E.*, Paper No. 1716 R, September, 1954 (**102** B, p. 544).

(20) Bramley, E. N., and Ross, W.: "Measurements of the Direction of Arrival of Short Radio Waves reflected at the Ionosphere," *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.

(21) "Radio Research 1949" (H.M. Stationery Office, Code No. 47–77–0), p. 30.

(22) Miya, K., Kobayashi, T., and Wakai, N.: "The Field Intensity of Scattered Wave in Radio Wave Propagation," Report of Ionosphere Research in Japan, 1951, **5**, p. 55.

(23) Taylor, D., and Westcott, C. H.: "Principles of Radar" (Cambridge University Press, 1948), pp. 44 and 135.

(24) Ridenour, L. N.: "Radar System Engineering" (M.I.T. Radiation Laboratory Series No. 1, McGraw-Hill), pp. 25 and 65.

(25) Eckersley, T. L.: "Studies in Radio Transmission," *Journal I.E.E.*, 1932, **71**, p. 405.

(26) Bremmer, H.: "Terrestrial Radio Waves" (Elsevier Publishing Co., 1949), pp. 168, 182 and 239.

(27) Shearman, E. D. R.: "A Study of Ionospheric Propagation by Means of Ground Back-Scatter" (see page 203).

(14) APPENDICES

(14.1) Equivalent Path/Ground Range Relations for a Curved Earth and Parabolic Layer

Appleton and Beynon[14] have considered ray propagation through a curved layer with a parabolic ionic-density/height relation. The various co-ordinates used by them are shown in Fig. 16. Their expression for the part ground-range D_1 corre-

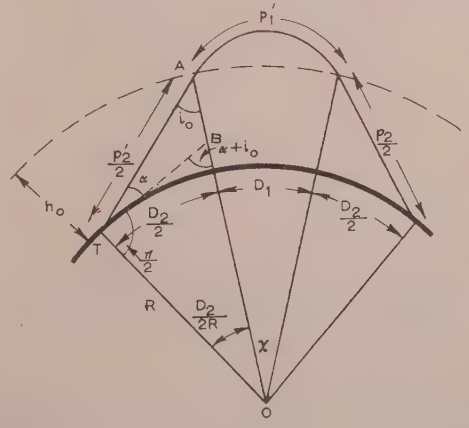


Fig. 16.—Co-ordinates of ray trajectory for curved earth and curved thick layer.

sponding to the path in the layer becomes, when the logarithm is put in the arc tanh form

$$D_1 = RX = \frac{R}{R+h_o} 2xy_m \sin i_o \operatorname{arc\,tanh}\left(\frac{x\cos i_o}{1-\frac{x^2y_m}{R+h_o}\sin^2 i_0}\right)$$

An argument closely resembling that used in the derivation of this equation leads to the expression for equivalent path in the layer

$$p'_1 = \int \frac{d_s}{\mu} = 2xy_m \operatorname{arc\,tanh}\left(\frac{x\cos i_o}{1-\frac{x^2y_m}{R+h_o}\sin^2 i_o}\right) = \frac{R+h_o}{R}\frac{D_1}{\sin i_o}$$

The path length below the layer and the corresponding ground range can be found by inspection of Fig. 16. TB is the tangent to the earth at T so that $\angle BTO = \pi/2$.

Therefore
$$\frac{D_2}{2R} + (\alpha + i_o) = \frac{\pi}{2}$$

$$D_2 = 2R\left(\frac{\pi}{2} - \alpha - i_o\right)$$

Considering the triangle AOT,

$$\frac{\frac{p'_2}{2}}{\sin\left(\frac{D_2}{2R}\right)} = \frac{R}{\sin i_o} = \frac{R+h_o}{\sin\left(\alpha+\frac{\pi}{2}\right)}$$

Hence
$$p'_2 = \frac{2R\sin\left(\frac{D_2}{2R}\right)}{\sin i_o} = \frac{2R\sin\left(\frac{\pi}{2}-\alpha-i_o\right)}{\sin i_o}$$

and
$$\alpha = \arccos\left(\frac{R+h_o}{R}\sin i_o\right)$$

The complete expressions for p' and D are thus

$$p' = p'_1 + p'_2 = 2xy_m \operatorname{arc\,tanh}\left(\frac{x\cos i_o}{1-\frac{x^2y_m}{R+h_o}\sin^2 i_o}\right) + \frac{2R\sin\left(\frac{\pi}{2}-\alpha-i_o\right)}{\sin i_o} \quad . \quad (11)$$

$$D = D_1 + D_2 = \frac{R}{R+h_o}\sin i_o p'_1 + 2R\left(\frac{\pi}{2}-\alpha-i_o\right) \quad . \quad (12)$$

These expressions have been used to calculate the curves of Figs. 9, 10 and 11. The procedure is to choose values of i_o, find α from $\alpha = \arccos\left(\frac{R+h_o}{R}\sin i_o\right)$, and then calculate p' and D from eqns. (11) and (12).

The expressions for p'_2 and D_2 give the total path length and ground range for reflection from a thin layer, and have been used to calculate the Es curve of Fig. 11.

(14.2) Calculation of Focusing Factor F_i/F_0

(14.2.1) Plane Earth.

Consider the energy in a ray pencil leaving a transmitter T (Fig. 12) bounded by rays at angles of elevation α, $\alpha + d\alpha$ and azimuthal angles θ, $\theta + d\theta$, θ_B in the figure being here replaced by $d\theta$. The area normal to this ray pencil at unit distance from the transmitter is $\cos\alpha d\theta d\alpha$, and the total energy flowing through it is $F_o \cos\alpha d\theta d\alpha$.

Let the change in ground range corresponding to $d\alpha$ be dD. The area normal to the same ray pencil at the point of incidence on the ground is $Dd\theta dD\sin\alpha$, and the energy passing through the area is $F_i Dd\theta dD\sin\alpha$. This energy can be equated to that at unit distance from the transmitter, whence

$$\frac{F_i}{F_o} = \frac{\cot\alpha}{D\frac{dD}{d\alpha}}$$

(14.2.2) Curved Earth.

The energy bounded by the same pencil of rays at unit distance from the transmitter is $F_o\cos\alpha d\theta d\alpha$, as before.

The area normal to the rays at the point of incidence on the ground is now (Fig. 15) $R\sin(D/R)d\theta dD\sin\alpha$, and the energy flowing through the area is $F_i R\sin(D/R)d\theta dD\sin\alpha$. Equating the two expressions for the energy in the ray pencil, we get

$$\frac{F_i}{F_o} = \frac{\cot\alpha}{R\sin\left(\frac{D}{R}\right)\frac{dD}{d\alpha}} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

In these expressions, D and R must be in the same units as those defining F_o. Thus if F_o is the power flux at 1 m, D and R must be in metres.

(14.2.3) Parabolic Layer.

Differentiating the expressions given in Section 14.1 for the ground range with a curved earth and curved, parabolic layer, we get

$$\frac{dD}{d\alpha} = \frac{dD_1}{d\alpha} + \frac{dD_2}{d\alpha} = \frac{R}{R+h_o}\frac{\sin\alpha}{\cos i_o}\left[2RA\frac{\left(1-\frac{2x^2y_m}{R+h_0}+A\right)}{(1-A)^2-(x\cos i_o)^2} - \cot i_o D_1\right] - 2R\left(1-\frac{R}{R+h_o}\frac{\sin\alpha}{\cos i}\right) \quad . \quad (14)$$

where $A = \frac{x^2y_m}{R+h_o}\sin^2 i_o$.

By substituting for $dD/d\alpha$ in eqn. (13) the focusing factor is obtained. This expression has been used to calculate the curves of Fig. 10 except at the skip distance, where the ray theory is not applicable.

(14.2.4) Incident Energy at the Skip Distance.

Bremmer[26] has given an expression for the field strength at the skip distance in terms of the second derivative of ground range with respect to the angle of elevation of the ray. His eqn. (42)* gives the radial component of the field set up at the

* Reference 26, p. 186.

kip distance by a short vertical transmitting aerial radiating 1 kW, taking into account the reflection coefficient of the ground at the transmitter and receiver.

Considering only one-hop propagation, Bremmer's expression, n the symbols used in the paper, becomes:

Radial component of field $E_r =$

$$\frac{0{\cdot}4886}{(\lambda_{\text{metres}})^{1/6}} \left| \frac{(1+\rho)^2 \cos^{5/2}\alpha}{\sin^{1/2}\left(\frac{D}{R}\right) \sin\alpha \left[\frac{\partial^2 \frac{D}{R}}{\partial(\cos\alpha)^2}\right]^{1/3}} \right| \text{mV/m} \qquad D = D \text{ skip}$$

where ρ is the reflection coefficient of the ground.

To find the total incident field, we put $E_i = \frac{E_r}{(1+\rho)\cos\alpha}$. For an isotropic radiator the field for a short vertical radiator must be multiplied by $\sqrt{\left(\frac{4}{3}\right)}\frac{1}{(1+\rho)\cos\alpha}$. Then, simplifying he derivative and remembering that $\frac{\partial D}{\partial \alpha} = 0$ at D skip,

$$E_i = \sqrt{\left(\frac{4}{3}\right)} \frac{0{\cdot}4886}{(\lambda_{\text{metres}})^{1/6}} \frac{\cos^{1/2}\alpha}{\sin^{1/2}\left(\frac{D}{R}\right)\left(\frac{\sin\alpha}{R}\frac{\partial^2 D}{\partial\alpha^2}\right)^{1/3}} \text{mV/m}$$

From the incident field the incident flux is found by $F_i = \frac{E_i^2}{10^6 \cdot 120\pi}$ watts/m². Then comparing F_i with the flux at 1 m from the 1 kW transmitter, $F_o = \frac{1\,000}{4\pi}$ watts/m², we get

$$\frac{F_i}{F_o} = 10{\cdot}59 \cdot 10^{-12} \frac{1}{(\lambda_{\text{metres}})^{1/3}} \frac{\cos\alpha}{\sin\left(\frac{D}{R}\right)} \frac{1}{\left(\frac{\sin\alpha}{R}\frac{\partial^2 D}{\partial\alpha^2}\right)^{2/3}} \qquad D = D \text{ skip}$$

This expression has been used to calculate the skip-distance values of $\frac{F_i}{F_o}$ in Fig. 10. $\frac{\partial^2 D}{\partial\alpha^2}$ has been found by plotting $\frac{dD}{d\alpha}$ obtained from eqn. (14) against α in the neighbourhood of the skip distance, and measuring the slope of the curve for $\frac{dD}{d\alpha} = 0$.

[The discussion on the above paper will be found on page 232.]

621.396.11 : 621.396.812.3

Paper No. 1901 R
Sept. 1955

AN EXPERIMENT TO TEST THE RECIPROCAL RADIO TRANSMISSION CONDITIONS OVER AN IONOSPHERIC PATH OF 740 KM

By R. W. MEADOWS, B.Sc.(Eng.), Associate Member.

(*The paper was first received 2nd March, and in revised form 21st May, 1955. It was published in September, 1955, and was read before the* RADIO AND TELECOMMUNICATION SECTION *31st October, 1955.*)

SUMMARY

Two-way pulse transmissions made simultaneously on the same frequency (5·1 Mc/s approximately) between Slough and Inverness (740 km) are described. An aerial common to transmitter and receiver was used at each terminal, and the fading patterns displayed at each end were compared visually: the Inverness display was relayed to Slough over a Post Office trunk line for this purpose.

The fading of corresponding echoes was found to be non-reciprocal for about 1% of the time during a total period of observation of about 15 daylight hours spread over 13 days in May, 1954. The precautions taken to ensure that the effects were ionospheric and not instrumental are outlined.

(1) INTRODUCTION

A radio path between a given pair of arbitrary aerials is here regarded as "reciprocal" if the ratio of signal power available from the receiving aerial to that fed into the transmitting aerial is the same, at a particular instant of time, whichever direction of transmission is used. Any divergences between the ratios obtained in the two directions would provide a measure of the non-reciprocity of the path. This definition follows from Carson's generalization[1] for radiating systems of Rayleigh's original reciprocity theorem,[2] and he showed that reciprocity is valid if the total current at any point (conduction plus displacement) is proportional to the electric intensity. In general, this is not so for an ionized medium in the presence of a steady magnetic field, and the ionosphere might be expected to cause differences between "go" and "return" paths in the following manner.

With no magnetic field present, a free electron in an ionized medium will be spun round by an incident circularly-polarized wave at such a radius, determined by the strength of the wave and the frequency, that the electric force balances the centripetal force on the electron.[3] This radius is independent of sense of rotation. When a steady longitudinal magnetic field is present, however, the radii for opposite directions of rotation are different, and the effect is that the right-hand wave (looking along the field) is propagated faster than the left-hand one, and this is true for either direction of wave propagation. The right-hand wave is the "extraordinary" component and the left-hand the "ordinary." The plane of the electric vector ("polarization") formed at any fixed point in the medium by a combination of two such contra-rotating waves of equal amplitude and frequency is then found to twist anti-clockwise with increasing distance measured along the direction of the longitudinal geomagnetic field, irrespective of the direction of propagation of the wave. The polarization is then non-reciprocal.

Consider, for example, a dipole turned through 45° with respect to another, both being at right angles to the line joining their centres, and spaced by the amount of loss-free ionosphere necessary to rotate the plane of polarization through 45°. The electric field from one dipole would then arrive parallel to the other for one direction of propagation and at right angles for the reciprocal direction. Evidently energy could then be received in one direction only. Budden[5] has considered propagation in the magnetic meridian generally (the example just cited is a case of this), and finds that between two horizontally or two vertically polarized aerials propagation should always be reciprocal. It should also be reciprocal in amplitude between one aerial of one kind and one of another, but with a phase reversal in one direction. A horizontally stratified ionosphere is assumed.

It might perhaps be thought that there was the possibility of the "go" and "return" paths being different; then, if the ionosphere were not homogeneous, absorption might be different in the two directions. However, thermodynamical considerations appear to indicate that the two paths would be identical, and as the attenuation per unit thickness of ionosphere is independent of direction for each magneto-ionic component, energy dissipation must also be independent of it. Booker[4] has also expressed ray characteristics at a point in terms of a parameter q (the vertical component of each vector refractive index) which is independent of direction, so that the paths traced out by component rays do not depend upon the latter.

The possibility of non-reciprocal effects has been realized since Carson's original paper[1] in 1924, and differences in performance in the two directions of transmission have often been observed. All such observations have, however, either been subjective, in which case noise might have modified the results, or have involved two-way comparisons of mean path attenuation; and this required accurate comparisons of transmitted and received power. It was felt that the existence of true non-reciprocity was still unproven, and an experiment not subject to possible errors of calibration was desirable.

This report describes such an experiment. Two-way pulse transmissions were carried out simultaneously at a frequency of 5·1 Mc/s between the D.S.I.R. stations at Slough and Inverness, which are 740 km (420 miles) apart. The direction of this path as seen from Slough is 8° west of the magnetic meridian.

(2) METHOD ADOPTED AND PRECAUTIONS NECESSARY

It was decided that a method which could demonstrate the existence of non-reciprocity qualitatively was called for, since, unless errors of measurement and sampling could be very accurately assessed, quantitative measurements might always be doubted. The most attractive method appeared to be to compare visually the instantaneous fading simultaneously at the two ends of the path. This implied that

(*a*) Pulse transmissions would have to be used, with pulse-echo amplitudes displayed on a linear time-base (A-display).

(*b*) An aerial common to transmitter and receiver would be required at each end to eliminate "space-diversity" effects; and a switch to connect it alternately to transmitter and receiver at the pulse-repetition frequency would be necessary, great care being taken to ensure that the reception and radiation characteristics remained identical.

(*c*) Care would have to be taken to ensure that the frequencies of

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

the reciprocal transmissions were sufficiently close for "frequency-diversity" to be negligible.

(*d*) A means for relaying the cathode-ray-tube display from one end to the other would be required in order that the two sets of pulse echoes could be observed together and compared.

(3) APPARATUS USED

(3.1) Inverness Terminal

An existing transmitter and aerial system was used. The transmitter delivered 150 microsec 5 kW pulses into a short 2-wire 600-ohm line feeding the base of a vertical 3-wire rhombic aerial. The rhombic was 85 ft high with side length of 46 ft and semi-side angle of 50° The plane of the aerial was at right angles to the Inverness–Slough meridian; radiation in the Slough direction was therefore horizontally polarized.

The balance of the peak voltage applied to the feeder could be checked by a high-impedance diode voltmeter. A switch comprising an artificial quarter-wave line and a pair of thyratrons in each limb of the line enabled the aerial to be connected alternately to transmitter and receiver at the pulse-repetition frequency of 50 c/s. A pulse receiver[6] of the type developed at the Radio Research Station, with built-in i.f. attenuator, was used to display the received echoes on a cathode-ray tube with linear time-base. Both transmitter and time-base were triggered 50 times per second by a Post Office quartz-crystal unit to a nominal accuracy of one part in 10^8. This precision was necessary to ensure that (with a similar unit at Slough) the two pulse displays would remain in substantially the same positions on the cathode-ray tubes for hours on end.

The A-display of signals received from Slough was scanned in the X-direction by a very narrow strobe, the output from which was held constant between successive pulses, so that a d.c. level proportional to gated A-display Y-voltage was produced; the maximum permissible scanning rate was 700 km of range per second. This d.c. level was used to amplitude-modulate an a.f. tone: another tone of different frequency was amplitude-modulated by a direct voltage defining the X-position being gated. These two tones were passed down a Post Office trunk telephone line to Slough, separated, amplified and rectified and applied to the Y- and X-plates, respectively, of a long-persistence cathode-ray tube. The Inverness display was thereby repeated at Slough as a succession of rapid "paints," and was available for comparison with the display received by direct ionospheric transmission; the delay introduced by the Post Office line was only about 100 millisec, and was therefore negligible compared with even the most rapid ionospheric single-ray fading periods.

(3.2) Slough Terminal

A single-ended transmitter delivered 150 microsec 5 kW pulses into a buried 75-ohm coaxial cable feeding a vertical 6-wire $\lambda/4$ unipole 45 ft high. Eight buried $\lambda/2$ radial earths were used. Radiation was therefore vertically polarized. Transmit-receive switching, receiver and pulsing arrangements were similar to those at Inverness.

(4) RESULTS

On the frequency of 5·1 Mc/s used, one-hop E and F (both ordinary ray) echoes were observable simultaneously on the 20 millisec time-base at each end, showing an effective path difference of about 100 km for most of the time. A convincing qualitative test of non-reciprocity would therefore have been obtained when the E-ray echo was bigger than the F in one direction and smaller in the other. A careful watch was therefore kept for this effect, which would have demonstrated the existence of non-reciprocity no matter how non-linear the receiving apparatus might have been. In practice, however, this was linear over a 25 dB range. A proper statistical analysis of amplitudes, although academically desirable, would have required considerable instrumental complication; continuous reliable logarithmic recording of gated E and F echoes closely synchronized between the two terminals would have been necessary.

The Inverness display was therefore adjusted so that only 1E* and adjacent 1F* rays were being scanned; about three "paints" per second were then possible, so that fading changes taking longer than $\frac{1}{3}$ sec could be observed at Slough. This was found to be adequate; and the duration of time during which non-reciprocity occurred was recorded.

Altogether, out of 886 min of observation during the hours of daylight on 13 days in May, 1954, effects which were definitely not reciprocal occurred for only 8·4 min, or 1% of the time. On three days the paths appeared entirely reciprocal; on the worst day violent non-reciprocity was apparent for at least 7% of the time. Also on this day the 1E ("ordinary") ray was sustained at Slough but disappeared completely (>25 dB change) at Inverness for about 10 sec.

Whenever noticeable effects occurred, the frequency of one transmitter was changed slightly in order to see whether they were due to frequency diversity caused by a slight misalignment of transmitter frequencies. On no occasion was this found to be so.

(5) DISCUSSION OF RESULTS

Evidently the paths were not always reciprocal, but so infrequently that measurements of amplitude averaged over short periods (say one minute) would have shown little significant difference in the two directions; in fact any differences which did occur could probably be explained by sampling or instrumental errors.

Little can be deduced about variation of long-term mean level between path attenuation in the two directions; the experiment was not designed with that object in mind. However, it seems most improbable that any substantial differences in the mean path attenuations would occur if so sensitive a test as fading showed little difference. This matter is being investigated in connection with 2-way absorption measurements made between Slough and Inverness during the solar eclipse of 30th June, 1954.

Bearing in mind the fact that only "ordinary" rays appeared to have been present in the measurements, the circuit should have been entirely reciprocal on the argument given in Section 1. However, it is possible that reflections from a sharp low-level boundary of ionization may have momentarily introduced some "extraordinary" component which, combined with the "ordinary" produced a plane of polarization different for the two directions of transmission. Unfortunately, calculation of these, and other non-reciprocal effects, is difficult, but would have to be attempted for a full understanding of the results to be obtained. Such calculations would probably also prove of great value in estimating m.u.f.'s and l.u.f.'s from known vertical-incidence data, but would be so cumbersome that they would have to be programmed for a computing machine.

(6) CONCLUSIONS

Differences between the fading patterns of single-hop ordinary E- and F-ray echoes observed on the same frequency (5·1 Mc/s) at each end of a 2-way ionospheric path between Slough and Inverness, a distance of 740 km, have shown that definite non-reciprocal effects occurred for about 1% of the time. It is considered from this result that the average path attenuation would not show any significant differences in the two directions. Steps taken (instrumentally) to ensure that the observed effects were in fact ionospheric have been outlined. Since non-

* The prefix 1 means "1-hop."

reciprocal amplitude effects have been shown to occur on this short path it may be expected that on longer-distance circuits more pronounced effects could be obtained. However, if these are in fact due to non-reciprocal swinging of the plane of polarization as has been suggested, they might still be expected to be of reasonably short duration, and to show little preference for a particular direction of transmission.

(7) ACKNOWLEDGMENTS

Thanks are due to the Post Office Engineering Department for providing the tone units used in relaying the Inverness display to Slough, and for organizing trunk telephone-circuit facilities. The author would also like to acknowledge most helpful discussions with Messrs. J. A. Ratcliffe, H. G. Booker and K. G. Budden, and with his colleagues at Slough, in particular Mr. W. R. Piggott.

The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(8) REFERENCES

(1) CARSON, J. R.: "A Generalization of the Reciprocal Theorem," *Bell System Technical Journal*, 1924, **3**, p. 393.
(2) LORD RAYLEIGH: "The Theory of Sound," Vol. 1, pp. 150–157.
(3) "Fundamental Principles of Ionospheric Transmission," D.S.I.R. Radio Research Special Report No. 17.
(4) BOOKER, H. G.: "Propagation of Wave-Packets incident obliquely upon a Stratified Doubly-Refracting Ionosphere," *Philosophical Transactions of the Royal Society*, 1938, **237**, p. 411.
(5) BUDDEN, K. G.: "A Reciprocity Theorem on the Propagation of Radio Waves via the Ionosphere," *Proceedings of the Cambridge Philosophical Society*, 1954, **50**, Part IV, p. 604.
(6) PIGGOTT, W. R.: "The D.S.I.R. Ionospheric Absorption Measuring Equipment," *Wireless Engineer*, 1955, **32**, p. 164.

[The discussion on the above paper will be found on page 232.]

621.396.2 : 621.396.11

Paper No. 1904 R
Oct. 1955

AN EXPERIMENTAL TEST OF RECIPROCAL TRANSMISSION OVER TWO LONG-DISTANCE HIGH-FREQUENCY RADIO CIRCUITS

By F. J. M. LAVER, B.Sc., and H. STANESBY, Members.

(*The paper was first received* 14*th April, and in revised form* 2*nd June*, 1955. *It was published in October*, 1955, *and was read before the* Radio and Telecommunication Section 31*st October*, 1955.)

SUMMARY

Tests have been carried out under carefully controlled conditions to see whether the attenuation of high-frequency signals sent over a given long-distance radio path differs according to the direction of transmission. The results obtained both across the North Atlantic and between Australia and the United Kingdom show that at times the loss in both directions is substantially the same, and that at other times the loss difference can rise to values of the order of 5 or 10 dB. Usually the loss differences were such that signals outwards from the United Kingdom suffered the greater attenuation, but without further evidence it would be unwise to assume that this tendency is permanent.

(1) INTRODUCTION

When considered in terms of the proportion of time suitable for commercial traffic, the performance of some of the long-distance high-frequency radio circuits operated from the United Kingdom is at times appreciably different for the two directions of transmission, due allowance being made for any differences in the power of the transmitters used. Furthermore, it has been noticed on some circuits that the direction of superior performance tends to change according to the season of the year. Why this should be is the subject of some speculation on the part of operating organizations.[1] It is possible that the characteristics of the aerials used for one direction of transmission sometimes happen to suit the propagation conditions of the particular radio path concerned better than do those for the reverse direction; it may be that the noise received at one end of the circuit is greater than that at the other end; perhaps there are differences in the design or operation of the equipment used for the two directions of transmission; and perhaps the staff at one receiving terminal demand better-quality signals before accepting traffic than do those at the other end. It would indeed be surprising if one or more of these factors were not to introduce a measure of asymmetry into the behaviour of most radio circuits. Nevertheless it is possible that another, more fundamental, factor is involved as well, namely that the attenuation of the radio signals in the ionosphere differs for the two directions of propagation.

It is known on theoretical grounds that the earth's magnetic field may influence radio-wave propagation through the ionosphere differently for opposite directions of transmission; but since it is not possible to deduce the extent of the differences to be expected in practice it is necessary to resort to experiment. No previous work on the subject could be traced, and so measurements have been made under carefully controlled conditions of the transmission losses for both directions of propagation over two different high-frequency radio circuits. The paths concerned were between the United Kingdom and the United States of America, and between the United Kingdom and Australia, and the frequencies used were of the order of 12 Mc/s.

Mr. Laver and Mr. Stanesby are in the Post Office Engineering Department.

(2) THE RECIPROCITY THEOREM AND HIGH-FREQUENCY RADIO CIRCUITS

In 1853 Helmholtz announced and proved a "reciprocal theorem" as an application to electrically-conducting bodies of an earlier theorem which Green had used for problems in electrostatics.[2] Lord Rayleigh later stated and proved the reciprocity theorem for alternating-current circuits;[3] and this was generalized and applied to extended electromagnetic systems by Carson,[4] who later stated the theorem in the more concrete form[5]

> If an electromotive force is inserted in the transmitting branch of antenna A_1 and the current is measured in the receiving branch of A_2, then an equal current (both as regards amplitude and phase) will be received in the transmitting branch of A_1 if the same electromotive force is inserted in the receiving branch of A_2.

In this form it is commonly called the Rayleigh–Carson Theorem.

Carson states[5] quite definitely that his theorem fails "when the waves are propagated in an ionized medium in which the earth's field has an appreciable effect on the conduction current," and notes that this makes its application to high-frequency transmission somewhat doubtful. Neiman[6] also considers that the reciprocity principle is inapplicable for radio-wave propagation through an ionized medium in the presence of an external magnetic field, but Ballantine had earlier commented[7] that, intuitive suspicions apart, this failure of reciprocity had not been definitely proved. It is believed that this statement, made in 1929, remained true until the measurements described in this paper were carried out.

When polarized light traverses a transparent medium in a direction parallel to an ambient magnetic field, its plane of polarization is rotated in a non-reciprocal manner. Lord Rayleigh used this phenomenon, known as the Faraday rotation, to produce a one-way optical transmission system. This extreme example of the non-reciprocal transmission of electromagnetic waves has recently been applied using ferrite materials in waveguides.[8] The possible contribution of the Faraday rotation to non-reciprocal ionospheric transmission is difficult to assess in view of the fact that, whatever the polarization of the transmitted signals, the received signals are elliptically polarized.

If, using isotropic aerials, the paths of lowest loss between the two termini differ for the two directions of transmission, in a practical test of reciprocity the directivity of the aerials used might cause differences in transmission loss. This could happen only if the refractive index of a given region of the ionosphere were different for waves travelling in opposite directions.

A difference in ionospheric absorption for oppositely-travelling waves would clearly lead to non-reciprocity, even with isotropic aerials, and might conceivably arise from differences in the size of the orbits followed by ionospheric electrons, leading to differences in collision losses.

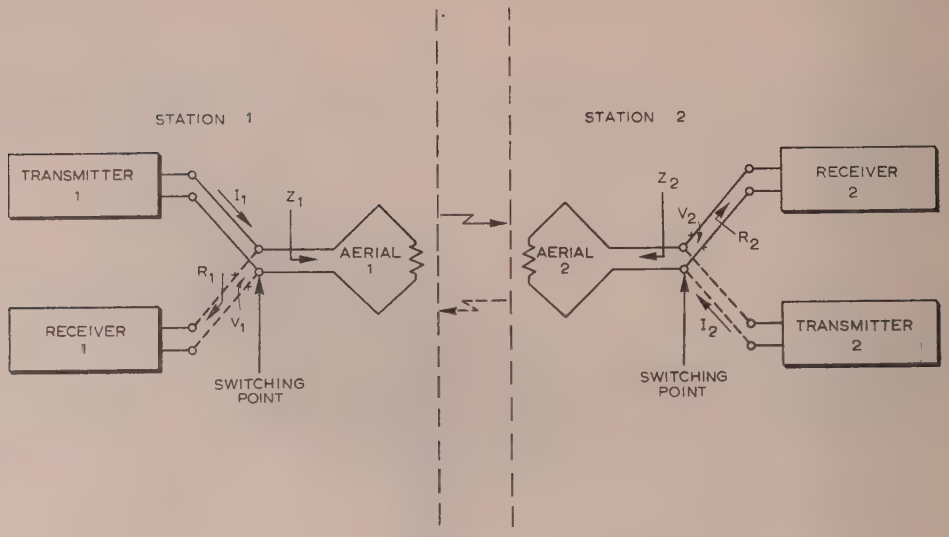


Fig. 1.—Test arrangement for radio-reciprocity tests.

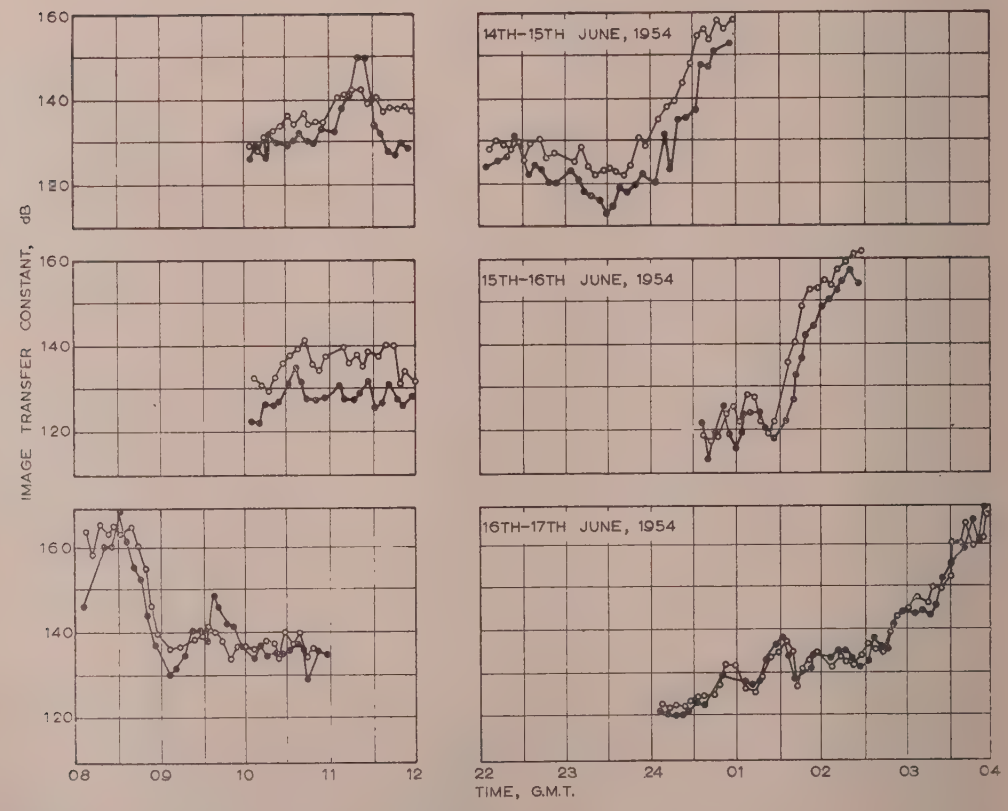


Fig. 2.—Measured image transfer constants for United Kingdom–United States North Atlantic Route, 13 332 kc/s, period 14th–17th June, 1954.

● United States → United Kingdom.
○ United Kingdom → United States.

(3) METHOD OF TEST

At each end of the radio circuit there was a single aerial and feeder which could be switched either to a transmitter or to a receiver tuned to the same frequency, as shown in Fig. 1. Transmissions of unmodulated carrier wave lasting for two minutes were made alternately from each end of the circuit with intervals of 30 sec between transmissions to allow for change-over. The current delivered by the transmitter to the aerial at one station and the potential difference developed across the input terminals of the receiver at the distant station were monitored. The output signals from the receivers were recorded, using time-constants of about 15 sec to smooth out the more rapid variations in signal strength, and the recorders were calibrated in terms of input-signal voltage. The results were analysed in the following way:

If, as shown in Fig. 1:

(*a*) The aerial and feeder impedance seen from the aerial-switching points is Z_1 ohms at Station 1, and Z_2 ohms at Station 2.
(*b*) The receiver input impedance seen from the aerial-switching point is R_1 ohms at Station 1, and R_2 ohms at Station 2.
(*c*) The current delivered by the transmitter to the aerial is I_1 amp at Station 1, and I_2 amp at station 2.
(*d*) The potential difference developed across the input to the receiver is V_1 volts at Station 1, and V_2 volts at Station 2.

it can be shown that the image transfer constant* of the path for transmissions from Station 1 to Station 2 is

$$20 \log_{10}\left[\frac{I_1}{V_2}\frac{\sqrt{(4Z_1Z_2)}}{(1+Z_2/R_2)}\right] \text{ decibels} \quad . \quad . \quad (1)$$

and for the reverse direction of transmission it is

$$20 \log_{10}\left[\frac{I_2}{V_1}\frac{\sqrt{(4Z_1Z_2)}}{(1+Z_1/R_1)}\right] \text{ decibels} \quad . \quad . \quad (2)$$

For reciprocity the two image transfer constants should be equal; hence, equating (1) and (2) gives

$$I_1V_1(1+Z_1/R_1) = I_2V_2(1+Z_2/R_2) \quad . \quad . \quad . \quad (3)$$

which resembles the form in which the reciprocal relation is expressed by Stevenson.[9]

This method of test by alternate transmissions in opposite directions is, of course, open to the possible objection that ionospheric changes between transmissions may confuse the results. Truly simultaneous and continuous transmissions on the same frequency would be virtually impossible in practice, and even if they were possible the finite time of transmission means that the received signal elements being examined at any instant at the two terminals could coincide only at one point in the path, and so very rapid ionospheric changes would introduce differences. However, since the practical interest lay in the occurrence and extent of non-reciprocal conditions persisting over relatively long periods, the method outlined above was considered to be satisfactory. Moreover, the extent to which the attenuations measured during successive transmissions in the same direction differ indicates the extent to which conditions are likely to have changed between transmissions in opposite directions, and this, it will be seen, is often quite small. No attempt was made to check reciprocity of phase, since the practical difficulties are so great and the practical interest so small.

* This quantity, more familiar in network theory than in ionospheric propagation studies, is used instead of insertion loss in order that the loss of the circuit can be expressed in terms that are not affected by the terminal impedances. The image transfer constant in decibels is $20 \log_{10}$ times the ratio of the input to the output in voltamperes when the network is operating between its image impedances. If for this condition the image impedances are equal, the image transfer constant gives the more familiar insertion loss.

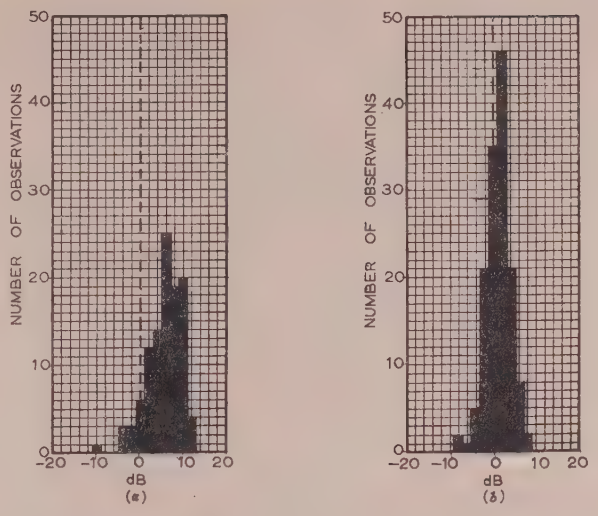


Fig. 3.—Histograms of image-transfer-constant differences for United Kingdom–United States North Atlantic Route, period 14th–17th June, 1954.

(*a*) 14th and 15th June, 1954.
(*b*) 16th and 17th June, 1954.

Note.—Positive values indicate United Kingdom → United States loss higher than United States → United Kingdom loss.

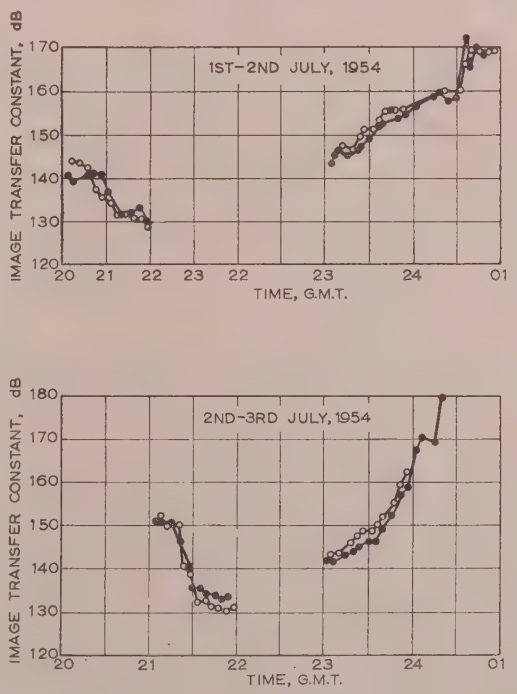


Fig. 4.—Measured image transfer constants for United Kingdom–Australia Asiatic Route, 11 992 kc/s, period 1st–3rd July, 1954.

● Australia → United Kingdom.
○ United Kingdom → Australia.

(4) TEST PROGRAMME

Tests between the United Kingdom and the United States were first made during two weeks in August and September, 1953, but interference from other transmissions largely spoilt the results. A more successful test series was completed in June,

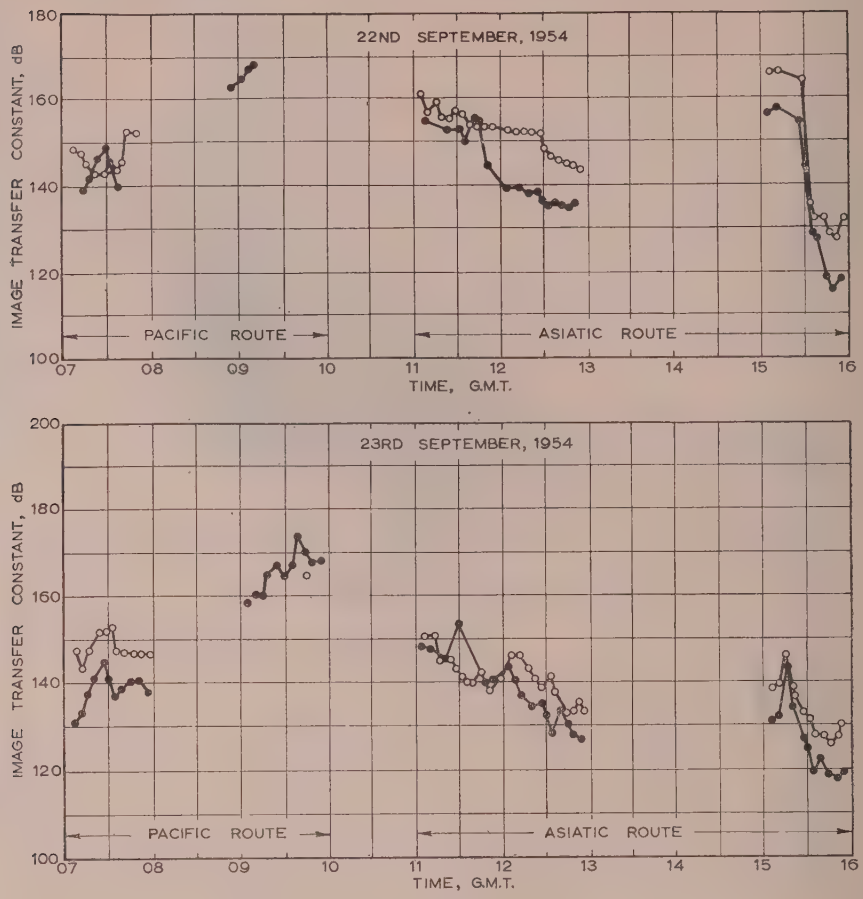


Fig. 5.—Measured image transfer constants for United Kingdom–Australia Asiatic and Pacific Routes, 11 992 kc/s, period 22nd–23rd September, 1954.

● Australia → United Kingdom.
○ United Kingdom → Australia.

1954, when just over 300 two-minute transmissions were satisfactorily received for each direction. The test frequency was 13 332 kc/s and each transmitter delivered some 500 watts to its aerial. A large rhombic aerial was used at Smallford, Herts, in the United Kingdom, and a Yagi array of about 10 dB gain was used at Murray Hill, New Jersey, in the United States. The image transfer constants were calculated by substitution in eqns. (1) and (2); in most cases Z_1, Z_2, R_1 and R_2 were about 600 ohms, I_1 and I_2 were between 0·7 and 2 amp, and V_1 and V_2 between about 3 and 100 μV. The measured image transfer constants for the two directions of transmission are shown plotted in Fig. 2 for typical test periods. Each plotted point represents an average taken over a 2 min transmission. The differences between the image transfer constants for the two directions of transmission have been evaluated by subtracting the mean of two adjacent United States → United Kingdom values from the intervening United Kingdom → United States value; i.e. a simple linear interpolation was made to offset the effects of systematic changes in loss caused by slow ionospheric changes. The differences for two typical sets of test periods, extending over 14th and 15th, and 16th and 17th June, have been grouped in 2 dB class-intervals, and are shown as histograms in Fig. 3.

Tests between the United Kingdom and Australia were made during five weeks in June, July and September, 1954. Rhombic aerials of similar dimensions were used at each end, namely at Smallford, United Kingdom, and at Doonside in New South Wales, Australia. Most of the tests were carried out on either 11 992 or 13 332 kc/s, although a few transmissions were exchanged on 14 300 kc/s. The majority of the transmissions were made over the Asiatic route, but some were sent over the Pacific route. In Figs. 4 and 5 the losses measured on 11 992 kc/s are shown for typical periods in July and September, 1954, respectively. Fig. 6 shows the differences for three sets of transmission periods plotted in histogram form.

(5) DISCUSSION OF RESULTS

The measurements made over the transatlantic path show that there are periods when transmission is substantially reciprocal within the accuracy of measurement, and others when it is persistently non-reciprocal. This is illustrated by comparing in Fig. 2 the results for 2400–0400 G.M.T. 16th–17th June, with those for 2200–0100 G.M.T. 14th–15th June, and by the histograms in Figs. 3(*a*) and 3(*b*). For the whole test period 14th–19th June, 1954, 272 image-transfer-constant differences were obtained. These results have a mean value of +3 dB in the sense that the United Kingdom → United States loss was greater than in the reverse direction, and they have a standard deviation

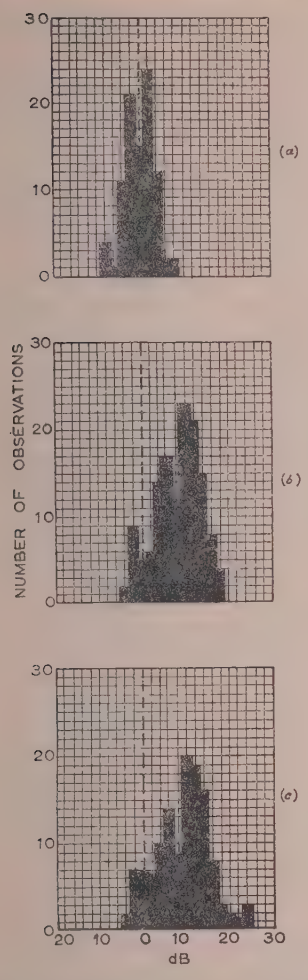


Fig. 6.—Histograms of image-transfer-constant differences for United Kingdom–Australia Asiatic Route.

(*a*) 11 992 kc/s, period 29th June–2nd July, 1954.
(*b*) 13 332 kc/s, period 6th–10th July, 1954.
(*c*) 11 992 kc/s, period 21st–24th September, 1954.

Note.—Positive values indicate United Kingdom → Australia loss higher than Australia → United Kingdom loss.

of 4 dB. Assuming that the results follow a normal distribution, the standard error of the mean would be about 0·25 dB, so that its 3 dB displacement from zero is significant. It does not necessarily follow that this 3 dB mean difference was due to ionospheric non-reciprocity; it might, for example, have been wholly or partly due to systematic differences between the transmitter current and receiver voltage standards used at the terminal stations. These have not been compared, but it is estimated that they might have contributed a standard error of about 1 dB to the measurements of image transfer constant, i.e. a standard error of about 1·5 dB to their differences. The most interesting result is not the 3 dB mean difference but the contrast between the more or less persistent difference in losses, amounting to 5 dB or more, observed on 14th, 15th and 16th June, and the near-equality observed on 17th June, and on the two following days, although to save space the last two days' results have not been reproduced. This change occurred without any alteration in test conditions—apart from those occurring in the ionosphere.

Again, the degree of non-reciprocity occasionally changed quite quickly, as in Fig. 2 between 1100 and 1200 on 14th June and between 0900 and 1000 on 16th June. These changes should, it is suggested, go far to allay any suspicion that the observed lack of reciprocity may have been due to instrumental errors.

The measurements made over the United Kingdom–Australia paths gave results largely similar to those for the transatlantic route. Thus, during the period 29th June to 3rd July, 1954, substantially reciprocal transmission was observed on 11 992 kc/s over the Asiatic route, as illustrated in Fig. 4. During this period 92 loss differences were obtained, and these have a mean value of 0 dB with a standard deviation of 3·5 dB, as shown in the histogram of Fig. 6(*a*). The same frequency and route were used for the test period 21st–24th September, 1954, but the 128 values of loss-difference obtained have a mean value of +9 dB, the loss in the United Kingdom → Australia direction predominating, and their standard deviation is 6 dB. Some typical results for this test period are shown in Fig. 5, and in histogram form in Fig. 6(*c*). It was possible to make only a few measurements over the Pacific route. A frequency of 11 992 kc/s was used and the results shown in Fig. 5 are generally similar to those for the Asiatic route.

Tests on 13 332 kc/s were made between the United Kingdom and Australia during the period 6th–10th July, 1954, and satisfactory results were obtained only for the Asiatic route. Substantially non-reciprocal conditions persisted during most of the test periods, as shown in the histogram of Fig. 6(*b*). The 135 values of loss-difference obtained were dispersed about a mean value of +8·5 dB with a standard deviation of 5·5 dB, the loss in the United Kingdom → Australia direction being higher, as before.

(6) CONCLUSION

The test results indicate that appreciable differences in transmission loss occur for high-frequency signals propagated in opposite directions over the same long-distance radio path. These differences appear to persist over periods of a few hours or a few days, and to occur quite frequently. The results obtained from a relatively short series of tests both on the North Atlantic and on the Australian routes, have yielded figures for attenuation that are, on the average, greater for outward transmission from the United Kingdom than in the reverse direction. Whether or not this would always be so is uncertain, but it is difficult to imagine why the long-term average of the transmission loss over any radio path should not be the same for both directions. It appears that there is room for considerably more experimental and theoretical work on this subject, which should commend itself to ionospheric physicists.

(7) ACKNOWLEDGMENTS

The equipment in the United States was established and operated by the Bell Telephone Laboratories and the American Telephone and Telegraph Company, and in Australia by the Overseas Telecommunications Commission (Australia). The authors acknowledge with grateful thanks the very full and willing co-operation of these organizations in the execution of the tests. The authors' thanks are also due to those of their colleagues who planned the tests in detail and carried them out, and to the Engineer-in-Chief of the General Post Office for permission to use the information contained in the paper.

(8) REFERENCES

(1) Humby, A. M., Minnis, C. M., and Hitchcock, R. J.: "Performance Characteristics of High-Frequency Radiotelegraph Circuits," *Proceedings I.E.E.*, Paper No. 1787 R, January, 1955, (**102** B, p. 513).

(2) G.W.O.H.: "The Application of Helmholtz's Theorems to Aerial Characteristics," *Wireless Engineer*, 1944, **21**, p. 153.

(3) LORD RAYLEIGH: "Theory of Sound" (Macmillan, London, 1894), **1**, p. 155 (2nd Edition).
(4) CARSON, J. R.: "A Generalization of the Reciprocity Theorem," *Bell System Technical Journal*, 1924, **3**, p. 393.
(5) CARSON, J. R.: "Reciprocal Theorems in Radio Communication," *Proceedings of the Institute of Radio Engineers*, 1929, **17**, p. 952.
(6) NEIMAN, M. S.: "The Principle of Reciprocity in Antenna Theory," *ibid.*, 1943, **31**, p. 666.
(7) BALLANTINE, S.: "Reciprocity in Electromagnetic, Mechanical, Acoustical and Interconnected Systems," *ibid.*, 1929, **17**, p. 929.
(8) HOGAN, C. L.: "The Ferromagnetic Faraday Effect at Microwave Frequencies and its Applications. The Microwave Gyrator," *Bell System Technical Journal*, 1952, **31**, p. 1.
(9) STEVENSON, A. F.: "Relations between the Transmitting and Receiving Properties of Antennas," *Quarterly of Applied Mathematics*, 1948, **5**, p. 369.

DISCUSSION ON THE ABOVE FOUR PAPERS BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 31ST OCTOBER, 1955

Mr. G. Millington: We have had two very interesting sets of papers on different aspects of long-distance high-frequency propagation, in which the authors have had the benefit of modern techniques. Mr. Shearman mentions the early work of Eckersley, who used a 40 kW transmitter with four fixed frequencies and an omni-directional aerial. Great progress has now been made by using a high-power rotating beam, a continuously variable frequency and a P.P.I. presentation.

Eckersley maintained that the scattering is from the E-layer, and although it has now been shown that most of the energy which returns to the transmitting point by reflection from the F-layer is scattered back from the ground, he was fundamentally right in saying that such scattering should occur. Has Mr. Shearman observed any signs of this back scatter from the E-layer ahead of the main ground-scatter return?

One would have expected that the ground scatter would have been much greater from mountains than from the sea, but there is no marked difference in the strength of the signals from the Atlantic and those scattered back to this country from other directions. Has the author any results on this aspect of the subject?

With regard to the papers by Messrs. Meadows and Laver and Stanesby, no mention is made of a letter to *Nature* describing an experiment carried out between Chelmsford and Bodmin by Eckersley *et al.** They had a variable-frequency pulse transmitter at each end, and showed that, as far as the maximum usable frequency was concerned, the circuit was reciprocal both for the ordinary and the extraordinary waves.

The magneto-ionic theory shows that reciprocity should hold so long as only one component is concerned. In general, both modes will be excited in the ionosphere by the incident wave, and the received reflected wave will contain a mixture of the two, unless the extraordinary wave is suppressed by absorption in the ionosphere or is separable from the ordinary echo by the use of pulses. I am convinced that the explanation of the apparent non-reciprocity in terms of two modes present, given by Mr. Meadows, is correct.

In the Post Office experiment we have a long-distance transmission using continuous waves, and the situation is much more complex than in the test described by Mr. Meadows. Not only is there more chance of obtaining a mixture of ordinary and extraordinary waves, but also the received signal is the resultant of a multi-path transmission. Even so, I feel that the non-reciprocity obtained must have another explanation than the failure of the magneto-ionic theory under conditions where reciprocity should apply.

Mr. C. M. Minnis: One of the activities at the Radio Research Station is to produce ionospheric data from which it is possible to calculate the maximum usable frequencies on radio circuits. It often happens that discrepancies occur when the calculated m.u.f.'s are compared with those actually observed on operational circuits; in particular, it is common to find that the observed m.u.f. exceeds that given by calculation. On a long-distance circuit there are many possible reasons for this, but it is not usually possible to obtain conclusive evidence as to which are the actual causes of the discrepancy. In contrast, the comparative simplicity of single-hop back-scatter propagation narrows down the range of potential causes of the differences between observed and calculated skip distances, and it seems that Mr. Shearman's work on these lines may have uncovered two sources of error in calculating m.u.f.'s.

The first concerns Fig. 5 of paper No. 1914 R by Mr. Shearman, in which the observed skip distance was found to be about 10% less than the calculated value. The basic ionospheric data were measured and not forecast, and it seems very unlikely that systematic errors in the vertical-incidence critical frequency can have been responsible; it is much more likely that the m.u.f. factor used in the calculation was wrong. The error could arise from the assumption that the distribution of electron density with height at the peak of the F2-layer can be represented by a parabola; although this is approximately true, it is usual to find that the parabola which fits best at heights very close to the peak of the layer is thicker than that at lower heights. A consequence of this is that, when the Appleton–Beynon and Booker–Seaton methods are used to calculate the true height of the F2-layer, the resulting value is too high, and because of this the calculated m.u.f. factor is too low. Errors in the m.u.f. factor of about −10% would be required to explain the discrepancies shown in Fig. 5, and the possibility just mentioned could easily account for these in many cases.

My second point refers to Fig. 2 in the same paper, in which the F1-layer appears to have been responsible for back-scatter signals at frequencies higher than those which could be propagated by the F2-layer. The F1-layer is usually disregarded in routine calculations of m.u.f., because it is assumed that any ray which penetrates the E-layer will also penetrate the F1-layer. Recent measurements on the F1-layer suggest that its true height is below rather than above 200 km, and it follows that the m.u.f. factors for the layer are higher than those normally used. This means that the m.u.f. for the F1-layer is considerably greater than that for the E-layer, and it is possible that the neglect of the F1-layer, particularly in summer, may account for some of the occasions when the m.u.f.'s on operational circuits are found to be well above those calculated using orthodox methods.

Mr. G. O. Evans: Some of the difficulties which arise in the planning of high-frequency radiocommunication are due to lack of direct measurements of the characteristics of oblique-incidence transmission. The advent of the techniques described by Mr. Shearman and the development of the oblique-incidence *p'f* equipment does, I think, represent a major advance in the methods

* ECKERSLEY, T. L., FALLOON, S., FARMER, F. T., and AGAR, W. O.: "Wireless Propagation and the Reciprocity Law," *Nature*, 1940, **145**, p. 222.

of investigating the properties of the ionosphere, especially those characteristics which are of importance to the communication engineer.

I am very interested to see the contrast between the comparative simplicity of the back-scatter records during night-time and during the winter months, and the increasing complexity of the records as summer is approached. Mr. Shearman raises the question of the practical significance of some of these complex modes of propagation in long-distance communication. I think they may possibly explain the reason for the fact that, whilst during winter we get fair agreement between the actual circuit performance and the performance predicted from calculation of m.u.f. and l.u.f., during summer daytime frequencies above the predicted m.u.f. can be used. I should like to know whether, as a result of these back-scatter experiments, we should pay more attention to the effects of various forms of intense E-layer ionization in preparing m.u.f. predictions.

The tests described by Messrs. Laver and Stanesby suggest that propagation is, if anything, more difficult in the direction from London to Australia. It is, however, a well-established fact that, from May to August, the circuit performance during the afternoon, when the Asiatic route is in use, is definitely worse in the direction incoming to London. A detailed examination of circuit data for the periods shown in Figs. 4 and 5 of the paper by Messrs. Laver and Stanesby shows that, in general, the day-to-day changes in image transfer constant are reflected in the circuit performance. This general agreement holds for transmissions both to and from London. Furthermore, the diurnal variations in image transfer constant are of the order we would expect from calculations of D-layer absorption.

From Fig. 2 of the same paper we find that for the New York–London telephone circuit the frequency-change time for 14 Mc/s agreed very closely with the time at which the image transfer constant reached a value of approximately 160 dB. It may, incidentally, be more than a coincidence that the reciprocal conditions on the morning of the 17th June were associated with very quiet magnetic conditions.

It thus appears that both for the Australian and New York tests the variations in propagation conditions shown by the authors' measurements are reflected more or less directly in circuit performance when each direction is considered separately. We might therefore expect the tendency for propagation conditions to be more difficult in directions outgoing from London to be shown in the overall circuit performance. The fact that, in the case of circuits to Australia, this is not so suggests that any differences in propagation conditions are outweighed by factors local to the receiving terminals. One of the more important of these factors is probably atmospheric noise, and measurements being made simultaneously at Somerton (England) and Rockbank (Australia) appear to indicate that, at the time of greatest difference in circuit performance, the atmospheric noise level at Somerton is at least 10 dB higher than at Rockbank. More detailed comment on these measurements must, however, await detailed examination of the results in Australia and in this country.

Mr. A. M. Humby: In the paper by Messrs. Laver and Stanesby mention is made of various factors which might be conducive to the asymmetry of performance with season in the two directions of a given radio circuit, to which attention was drawn in a paper by Messrs. Humby, Minnis and Hitchcock.* As pointed out in the latter paper, however, the asymmetry is observed to commence and end at similar times and seasons by two different organizations operating on substantially similar radio paths. For these two organizations (one military and the other civil) there inevitably exist differences in sites, continuity of skilled personnel, keying systems, design and operation of equipment, frequency usage, quality of signal demanded, and even differences in traffic loading. It seems most unlikely, therefore, that the asymmetry to which we have referred can be attributed to any significant extent to terminal differences of this type. Since the publication of this paper we have been able to show (see Table A) that, for

Table A

Location of transmitter	Bearing from London (E. of true North)	Time of onset of asymmetry
		hours G.M.T.
Harman (Canberra) ..	65°	0930
Singapore	78°	1330
Colombo	93°	1430
Simonstown (Cape Town)	164°	1730

reception in the United Kingdom in summer of signals from various directions, the time of onset of asymmetry becomes progressively later as the azimuthal bearing of the incoming signals increases. It was suggested in the paper that atmospheric noise might well be the main cause of the asymmetry of performance. It can now be shown that the major contributory factor is almost certainly an increase in the noise/signal ratio; a condition which is believed to be associated with the westerly movement—with the sun—of regions of high thunderstorm activity.

Mr. W. R. Piggott: I should like to ask Messrs. Laver and Stanesby about the absolute accuracy of their measurements, as it is important to know whether the patterns which they have shown really indicate that the loss in one direction is always greater than that in the other, or whether a systematic error is possible. It is notoriously difficult to remove all sources of systematic error from measurements of this type.

I should also like to make two comments on questions which other speakers have raised. The first is on scatter from the ground and from oceans, and whether one ought to see mountain ranges. An important point in Mr. Shearman's paper, and one which we always have to remember, is that the ordinary countryside is quite rough from the radio point of view. If we consider a big mountain, the mountain itself is still rough in the Rayleigh sense and is not necessarily a better echoing object than the ground as a whole.

If we consider this argument in more detail, we find that the only conditions under which it is possible to get an enhanced echo are those for which the phase of the reflected wave is held constant over a wide area, so that we have coherent mirror-type reflection. In practice the only cases which have been reported of abnormal back scatter from distant mountain masses fall into this class. For example, the ionospheric station at Lindau, Harz, in Germany, is almost at the centre of curvature of the Alps, and frequently receives an abnormal echo coming from that arc. The beginning of the Alps forms a sort of curved mirror, and while the mountains themselves do not contribute much, the gradual rise at their beginning does.

The other two examples, i.e. reflections from the Himalayas and from the Rocky Mountains, are also cases where the echo pattern appears to come from the initial rise of a mountain mass, this rise forming a more or less regular arc across a comparatively flat plain.

It is worth noting that waves comparable with Mr. Shearman's scattering sources would be common for average sea conditions, and a serious decrease in scattering would be expected only for exceptionally calm conditions.

* Humby, A. M., Minnis, C. M., and Hitchcock, R. J.: "Performance Characteristics of High-Frequency Radiotelegraph Circuits," *Proceedings I.E.E.*, Paper No. 1787 R, January, 1955 (102 B, p. 513).

I should also like to make a comment on the m.u.f. discrepancies referred to in Mr. Shearman's paper. I think that Mr. Minnis's suggestion has some truth, but it is worth considering that there are other sources of error present whenever one makes an ionospheric measurement which will cause the m.u.f. factor to be too small by an amount which will increase with distance.

Thus all conventional *h'f* recorders give values of the apparent height which are systematically too high by 5–15 km, and this introduces a small but significant consistent error in the m.u.f. factor deduced.

In addition, all practical m.u.f.-factor computations are based on "no field" theories of analysis of the layers and of propagation. When we include the magnetic field, two main sources of error are disclosed, first that more ionization is required to reflect the oblique wave than would be expected, and secondly that the apparent thickness of the reflecting layer is usually much greater than the real thickness. At short ranges there is no net error, but at long ranges the latter effect predominates and the m.u.f. is greater than would be calculated from a "no field" analysis.

Thus the obvious corrections to the simple theory both act in the same way and together give a systematic error which should be observable. We may conclude from Mr. Shearman's work that these errors are present and that it might be worth while to modify the prediction systems so as to remove them.

Dr. R. L. Smith-Rose: Reciprocity of transmission has been concerning those involved in radio research for the last thirty or forty years, and it is intriguing to find that we are to-day trying to discover whether or not reciprocity is strictly applicable to radio transmission over 100 years after the principle was first enunciated by Helmholtz, and over 60 years since it was advanced by Lord Rayleigh.

It is very difficult to discover whether, over any circuit from here to Scotland or to the United States or Australia, it is possible in any space of time to send more words in one direction than can be received in the reverse direction. This type of investigation is generally confused by differences in the performance of the equipment, difference in the skill of the operators, and the bugbear of local interference due to noise or any other source. Even to-day it is not easy to separate actual radio transmission conditions from these other effects, and it would therefore appear that there is still room for really crucial experiments in order to check the conditions under which reciprocity is applicable. I think that these experiments must be done by those who are concerned with investigating the ionosphere, since it is partly a problem of physics and partly a problem of communication practice.

The question of reciprocity has occurred in a number of applications, and I remember a case in direction-finding when, instead of determining directions by rotating a loop at the receiver and taking bearings on an instrument from an open-aerial transmitter, a system was evolved in which a loop was rotated at the transmitter and bearings were obtained by a timing process and observing the signals on an open-aerial receiver. A knowledge of the reciprocity theorem enabled us to predict more or less what the measured errors would be at the site at which the transmitter was installed, together with errors that might be caused owing to ionospheric transmission effects.

Mr. J. A. Smale: In the early days of long-wave working across the Atlantic there was presumably even less likelihood of non-reciprocal or diurnal variations of signal strength, and yet the hours of satisfactory working, apart from being limited, occurred at different times in the two directions. This could only be due to non-reciprocal noise levels, and it is therefore gratifying to learn that full-scale comparisons of noise at the two ends of a long-distance circuit are being carried out in a properly controlled manner. At present we do not know what degree of difference of signal/noise ratios is causing the observed degree of non-reciprocal circuit working. Some years ago, when testing the relative performance of directive receiving aerials on the Melbourne–London beam circuit, we found that an improvement of 2 dB in signal/noise ratio resulted in an increase of five hours a day in circuit working time. A general deduction from these figures may not be permissible, but it may indicate that the differences we are looking for are quite small. On the London–Capetown beam circuit there were non-reciprocal working periods which reversed with the seasons and could be explained by the movement north and south of the main noise source in the tropical areas on the route between the terminals. But one observed effect of non-reciprocal conditions is not so easily explained by different noise levels. This is the transmission of 11 Mc/s signals from Melbourne over the short great-circle paths to London and on to the Barbados relay-station situation on the same great circle. In certain periods of the sunspot cycle, in autumn (northern hemisphere), signals faded out at London at about 2000 hours G.M.T. but continued to be received at Barbados for some two hours and were relayed back to London in the same 11 Mc/s band. This certainly points to something non-reciprocal in the ionosphere.

The authors state that signal-strength recorders had a 15 sec smoothing circuit which removed the effects of rapid fading. Telegraph circuits are very susceptible to this type of fading, and it may be that it could be non-reciprocal owing to some ionospheric effect. Mr. Meadows states that on one of three days the Slough–Inverness circuits were non-reciprocal for 7% of the time. This is an important difference, but it is not clear in what way the day differed from the two others.

Mr. P. G. Redgment: There is another aspect of ground scatter which I feel deserves some attention. In direction-finding it has been noted that errors can be caused by scattered radiation reaching the instrument from random directions; when assessing the relative merits of d.f. sites it would be most useful to be able to estimate the probable amplitude of such scattered signals under any given propagation conditions, and it seems that Mr. Shearman's results may assist in this. It has been pointed out that, in fact, apparently smooth ground is probably rough to radio waves. On inspection, however, there would seem to be a big difference between broken country—not necessarily mountainous, but country which has suffered recent tectonic disturbance, and which appears to be a mass of corner reflectors—and the smooth rounded surface of waves in the open ocean, even with the crests taken off by wind. From normal radar experience it would seem reasonable to expect a very substantial difference in the amount of scattered energy from those surfaces, and if there is any experimental or theoretical evidence for or against this view it would, I believe, be interesting to all engaged in direction-finding. Any data on the actual scattering power of particular types of surface would also be useful.

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Mr. E. D. R. Shearman (*in reply*): Mr. Millington mentioned that Eckersley was fundamentally right in suggesting that long-range back-scatter from clouds and irregularities in the E-region should be possible. The failure to detect such echoes is presumably due to the use of insufficient transmitter power, since short-range direct echoes from the E-region have been seen, although they are of considerably lower amplitude than meteor echoes. As mentioned in the papers, low-intensity meteor

echoes are occasionally seen at a slightly shorter range than ground back-scatter, and with higher power, therefore, E-region scatter should also be detectable. In a recent paper* Silberstein has explained some discrepancies in his observations by postulating long-range E-region scatter, but it is not certain that this is the correct explanation of his results.

Referring to the question of the relative amplitude of scatter from land and sea, no further evidence is yet available, and the need for a thorough investigation is clear. The direct measurement from an aircraft of the scattering properties of various kinds of terrain at these wavelengths would provide much useful evidence.

Mr. Piggott's remarks about the influence of the position of the transmitter in relation to a mountain range are of interest, since they provide a possible explanation of the reception of echoes from the Alps at Lindau and not at Slough.

Mr. Redgement's comments on the corner reflectors found in terrain subject to recent geological disturbance suggest a further mechanism by which enhanced echoes could be returned. The reduced echo strength expected from a part of the sea subject to a dead calm should be detectable by systematic observations of back-scatter intensity.

One very important potential application of back scatter should be mentioned for which a knowledge of the scattering power of various forms of terrain would be necessary. This is the determination, from the intensity of the scattered echoes returned from a distant point, of the field strength existing at that point. Such a measurement should extend the practical application of the technique to greater ranges, and would provide a very valuable tool to the communicator, particularly for broadcast and mobile services.

I was very interested to learn that Mr. Evans has found that, whereas in the winter there is good agreement between predicted and observed m.u.f.'s, in summer daytime, frequencies higher than the predicted m.u.f.'s can be used. The back-scatter patterns certainly indicate that modes of propagation involving Es-reflection are present, and further studies may show how to allow for these in predicting m.u.f.'s.

Mr. R. W. Meadows (*in reply*): I infer from Mr. Smale's remarks that marginal differences (say 2 or 3 dB) in the signal/noise ratios at the terminals would often be sufficient to account for quite pronounced asymmetrical traffic effects on telegraph circuits operated under difficult conditions, since presumably the performance of teleprinters and undulators worsens rather suddenly when a minimum signal/noise ratio is reached. If this is so, asymmetrical effects must always, surely, be expected on all but the least troublesome circuits, since it is too much to hope that operating conditions at the terminals can be equalized so closely that transmissions in the two directions always become unworkable simultaneously. In this context Mr. Humby's remarks on noise levels show the kind of thing to be contended with. Also, had the 740 km Slough–Inverness path been used for a commercial circuit, the Slough-to-Inverness traffic would undoubtedly have proved more workable than that in the reverse direction, since the noise level at the Slough receiver was consistently some 15 dB higher than that at Inverness.

I feel, moreover, that true non-reciprocity in the ionosphere is not usually to blame for asymmetrical traffic effects, since most of those described seem to favour either one terminal or the other for considerable periods at a stretch. Ionospheric non-reciprocity should, on the other hand, appear as a kind of polarization fading caused (as Mr. Millington confirms) by interaction between ordinary and extraordinary components, and would not be expected to favour one direction more than the other. This was, in fact, usually observed in the fading of the echoes on those occasions when non-reciprocal effects were apparent in the tests carried out at Slough and Inverness, and this included the 7% day taken up by Mr. Smale; on this day the complete disappearance of the echo in one direction is not considered significant, since continuation of the measurements would undoubtedly have shown this rather rare effect to have occurred for the opposite direction also. No correlation with ionospheric storminess was observed on that day, although, in general, increased activity might tend to cause non-reciprocal effects if the distribution of ionization were altered so as to increase the possibility of the extraordinary component. I would, in this sense, tend to agree with Mr. Evans's implication of correlation with magnetic activity, although changes of ionization which might be so caused would also tend to affect the level of noise and interference at the terminals. Incidentally, I am sure Mr. Smale would agree that his Melbourne–Barbados–London phenomenon is not a violation of reciprocity as defined in both papers, since it appears to be explicable in terms of skip distances and variation of maximum usable frequency along the path.

Messrs. F. J. M. Laver and **H. Stanesby** (*in reply*): We are interested to note Mr. Millington's view that non-reciprocity requires the simultaneous presence of two or more modes of propagation: this condition is, of course, often satisfied by working high-frequency circuits. The purpose of our tests was not to check the magneto-ionic theory as such, but to evaluate empirically the approximate magnitude of any non-reciprocal propagation effects. We are obliged to Mr. Millington for drawing our attention to the work by Eckersley and others.

Mr. Evans comments that the measured path losses show that transmission should be more difficult from England to Australia, whereas the traffic data show the opposite. In the Introduction to our paper we listed some of the many factors that can contribute to asymmetry of radio-circuit performance, and which, in a particular case, might easily obscure the effect of the relatively small observed differences in transmission loss. Among the items listed was the difference in the noise power received at the two terminals, and it is interesting to see that Messrs. Evans, Humby and Smale all refer particularly to this factor. Mr. Smale's comment that small changes in signal/noise ratio can lead to large changes in performance is particularly true of telegraph systems, and especially of those using frequency modulation when they are operating with weak signals near the improvement-threshold value.

Mr. Piggott rightly questions the absolute accuracy of measurement, but in Section 5 of the paper we refer to this point and estimate that systematic errors might contribute a standard error of about 1·5 db. The most significant feature of the results appears to be not the absolute values of the observed mean differences in path attenuation but the changes between periods of near equality, e.g. Fig. 2, 16–17th June, and those of persistent difference, e.g. Fig. 2, 14–15th June, without any change in test conditions—other than ionospheric changes.

We agree with Dr. Smith-Rose that the study of reciprocity is proper to those already studying the ionosphere, and we shall watch for further results with interest.

The subject of reciprocity is a strange one for we tend to feel, almost instinctively, that reciprocity is right and natural, and to look for alternative explanations when it is called into question. It is therefore worth remembering that not all natural phenomena are reciprocal, as the Faraday rotation plainly shows.

* Silberstein, R.: 'Sweep Frequency Back-scatter—Some Observations and Deductions,' *Transactions of the Institute of Radio Engineers*, Vol. AP-2, No. 2, April, 1954.

621.396.11

Paper No. 1920 R
Oct. 1955

V.H.F. PROPAGATION BY IONOSPHERIC SCATTERING AND ITS APPLICATION TO LONG-DISTANCE COMMUNICATION

By W. J. BRAY, M.Sc.(Eng.), J. A. SAXTON, D.Sc., Ph.D., Members, R. W. WHITE, B.Sc., F.Inst.P., and G. W. LUSCOMBE, B.Sc.(Eng.), Associate Members.

(*The paper was first received* 16*th April, and in revised form* 6*th July*, 1955. *It was published in October*, 1955, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 31*st October*, 1955, *and the* CAMBRIDGE RADIO GROUP 17*th January*, 1956.)

SUMMARY

The paper describes an investigation of the propagation of v.h.f. radio waves by scattering from the E-region of the ionosphere. The dependence of the characteristics of the received signal on frequency, distance and aerial directivity is examined, and observations are made on the diurnal and seasonal variations of the received signal strength; the results are discussed in relation to existing theories.

Investigations are also described on the suitability of this form of propagation for the transmission of frequency-shift telegraphy signals, and for telephony signals transmitted by single-sideband amplitude modulation, and by phase modulation, of a carrier. A form of wave-angle diversity reception which yields a substantial improvement in the performance of a telegraphy system is described. The possible applications of v.h.f. scatter propagation to commercial services are discussed.

LIST OF SYMBOLS

$\sigma(\theta, \chi)$ = Scattered power per unit incident power flux density, per unit macroscopic element of volume of the scattering region (i.e. large compared with l^3) and per unit solid angle in a direction making an angle θ with the direction of the incident radiation.

l = Scale of turbulence.

θ = Scattering angle, i.e. angle between directions of incident and scattered radiation.

χ = Angle between the electric vector in the incident wave and the direction of scattering.

ϵ = Permittivity of scattering medium.

$\overline{\left(\frac{\Delta\epsilon}{\epsilon}\right)^2}$ = Mean square fractional variation of permittivity due to turbulence.

N = Mean electron density.

$\overline{\left(\frac{\Delta N}{N}\right)^2}$ = Mean square fractional variation of electron density.

f = Radio frequency.

λ = Radio wavelength.

f_N = Plasma frequency, i.e. critical frequency at normal incidence for a layer of electron density N.

λ_N = Plasma wavelength corresponding to N.

P_t = Transmitted power.

P_r = Received power.

P_r/P_t = Transmission loss.

D = Transmission distance.

A_t, A_r = Effective apertures of transmitting and receiving aerials (= A when identical).

G_t, G_r = Transmitting and receiving aerial gains relative to an isotropic aerial.

b = Thickness of the turbulent scattering layer.

n, p, q, r, s = Exponents of G_t, G_r, $\sin\theta/2$, f_N and f respectively in the transmission-loss equation.

β = Semi-angle of the radiated beam.

α_t, α_r = Cross-sectional areas of the radiated and receiving aerial beams respectively at the mid-point of the scattering volume.

V = Effective scattering volume.

Mr. Bray and Mr. White are in the Post Office Engineering Department, Radio Branch.

Dr. Saxton and Mr. Luscombe are at the Radio Research Station, Department of Scientific and Industrial Research.

(1) INTRODUCTION

It is now some 25 years since it was first pointed out by T. L. Eckersley,[1,2] as a result of his observations on short-wave scattering in the ionosphere, that the E-layer is a complicated structure of ionic clouds. Eckersley observed that scattering due to such clouds is significant at all wavelengths from about 50 m downwards; it is probable, however, that some lower limit of wavelength was envisaged, but not then specified. Eckersley further concluded that the ionic clouds must have some degree of permanence, since they would otherwise rapidly diffuse into one uniform layer; he also deduced that there must be a limit to the height of the scattering region, since the time of diffusion, which depends upon the pressure and decreases rapidly with height, is such that significant scattering would not be expected at heights much greater than about 115 km. The early work of Ratcliffe and Pawsey[3] on the correlation of the fading of high-frequency transmissions observed with spaced receivers also showed that the distribution of ionization in the E-region must be of an irregular nature and such as would produce incoherent scattering.

In further studies[4] of short-wave transmission Eckersley suggested that the E-region must be in a continuous state of turbulence and that the ionic clouds are being formed continually. He went on to develop a theory of the scattering process on the basis of spherically symmetrical clouds, assumed to have a radius large compared with the wavelength, and showed that, ignoring any effects of absorption (which would become more important as the wavelength increases), such clouds would be more efficient scatterers at the longer wavelengths, in accordance with his experimental observations in the band 14–50 m. Eckersley's investigations related mainly to back-scattered radiation, i.e. with the receiver near to the transmitting station, or else to forward scattering at large angles over relatively short distances (of the order of 20 km); but he certainly appreciated that scattering would be a major factor in practically all short-wave transmissions, and that it could vitally affect direction-finding.[5]

An important development in the study of E-region scattering took place in 1952, when Bailey and his colleagues[6] described an investigation in which the continuous transmission of a 49·8 Mc/s signal had been observed over a path of 1 245 km between Cedar Rapids and Sterling, U.S.A. The signal, the amplitude of which varied rapidly in a manner representative of a Rayleigh distribution, was found to be present irrespective of

time of day and season, and of geomagnetic disturbance. Although it was appreciated that meteoric ionization probably makes a contribution to the received signal, the suggested explanation of the observed phenomena was based on an extension of the theory of scattering due to turbulence in the troposphere. Scattering in the lower atmosphere had previously been discussed by Booker and Gordon[7] and by Megaw[8,9]; it was the theoretical formulation due to Booker and Gordon which was extended by them to the ionospheric case, it being assumed that the turbulence was that resulting from winds known to be present in the E-region. It thus seems clear that Bailey and his co-workers attach rather more importance to these turbulent fluctuations of electron density than to those, also occurring in the E-region, which must be associated with the production of ionized meteor trails. On the other hand, there is a school of thought[8,9,12] which considers that meteoric ionization is the dominant, if not the only, cause of long-range (up to about 2000 km) v.h.f. transmission via the ionosphere. It may well be found ultimately that the v.h.f. transmission phenomena observed are due to the effects of more than one agency. Isted,[13] for example, considers that atmospheric electricity, and particularly electrical discharges not necessarily accompanied by lightning, is mainly responsible for the scattering described by Eckersley. He further suggests that partial ionization of the E-region by conduction currents of atmospheric electricity is possible, and that this is probably one of the main mechanisms by which v.h.f. transmission over long distances by means of the ionosphere is supported. The present authors, however, are inclined to the view that the explanation of the characteristics of this type of propagation is primarily to be sought in terms of scattering due to turbulence, and from meteor trails, the relative importance of the two mechanisms not yet being clear.

The turbulence theory[6] based on the work of Booker and Gordon[7] and the treatment of the same problem by Villars and Weisskopf[14] on a somewhat different basis, both assume that the turbulent fluctuations of refractive index are independent of direction. On the other hand, there is, in general, a cylindrical symmetry of the ionization density in meteor trails, and the orientation of a given trail in relation to the transmission path will obviously be of importance; it might thus be expected that there would be some difference between the predictions of the two theories concerning the dependence of the properties of the scattered signal on frequency and aerial characteristics; that this is so is shown by the work of Eshleman and Manning[10] and McKinley.[11] As will be seen from Section 13.1, however, the predictions of the two theories are not so dissimilar that it is possible easily to distinguish the relative significance of the two scattering mechanisms without carrying out further experiments designed specifically to test critical points of difference between them.

The general aim of the theoretical work has been to determine the ratio of the received power to the transmitted power over a given transmission path, for a given wavelength and aerials of a given gain; as shown in Section 13.1, all that can be expected is approximate conformity with a relation such as

$$\frac{P_r}{P_t} \propto \frac{G_t^n G_r^p f_N^r}{f^s (\sin\theta/2)^q} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

A discussion of possible values for the various exponents is given in Section 13.1; it would appear on theoretical grounds that n and p both lie in the range 0–1 but are not necessarily equal; r is 4 for turbulent scattering, and may be expected to be greater than 0 for meteoric scattering, although its precise value is unknown; s will probably lie in the range 2–6, and q in the range 1–5, although there may be conditions under which even higher exponents might occur.

The work of Bailey and his colleagues[6] is of great interest to engineers concerned with the provision of long-distance point-to-point radio links, since it demonstrates that under suitable conditions scatter signals are present continuously. Moreover, such signals appear to be uninterrupted, and indeed are often enhanced, by ionospheric disturbances which cause considerable disruption to normal short-wave services. These facts, and the possibility of utilizing a region of the spectrum hitherto unexploited for long-distance radio links, led to a decision by the Post Office to investigate the characteristics and possibilities of v.h.f. propagation by ionospheric scattering at first hand. Accordingly, in 1952, an experimental link was set up between a transmitting station in the Shetland Isles and a receiving station at Jersey in the Channel Islands (Fig. 1), a distance of 1185 km (740 miles). Since, as explained above, the subject has considerable scientific interest, the Department of Scientific and Industrial Research also decided to co-operate in the investigation, and a second receiving station was set up at the Radio Research Station, Slough, 935 km (580 miles) from Shetland.

(2) HISTORY AND SCOPE OF THE INVESTIGATIONS

From an engineering aspect the investigation had two major objectives. First, it was desired to assess the extent to which ionospheric-scatter propagation might be used in the establishment of commercial telephony or telegraphy circuits, and to obtain data for the design of such systems. Secondly, it was required to assess the extent to which distant high-power transmitters might interfere with other services as a result of this type of propagation.

The scientific aspects of the investigation were primarily concerned with assessing the validity of the various theoretical hypotheses discussed in Section 1 and Section 13.1.

(2.1) Choice of Paths

It was apparent at the outset of this work that a transmitter of very high effective power would be required, and consideration was first given to the possibility of using a v.h.f. transmitter already in regular operation in Western Europe. The transmission which appeared to be potentially most useful was the amplitude-modulated 93·8 Mc/s signal from the B.B.C. experimental station at Wrotham. This has an effective radiated power in the horizontal direction of some 120–150 kW, but most of the energy is radiated at too low an angle for efficient transmission via the E-layer, for paths of a length practicable within the British Isles. An attempt was made to receive the Wrotham transmission in Shetland using a rhombic aerial and a high-sensitivity receiver; but over the period of the tests signals were received for periods of only a few seconds at a time and there was no trace of any residual or background signal.

It was concluded that a special transmitting station would have to be established. Such a station could make use of a high-gain aerial system in order to reduce the transmitter power required for a detectable received signal, and could be operated for 24 hours per day when required, in order to determine the diurnal variations of the received signal level. Modulation tests could also be made without interruption of the normal schedule of an operational station.

Having decided on the use of a special transmitter, it now remained to select the path, and one from the Shetland Isles to the Channel Islands was chosen as being the longest practicable within the British Isles. The transmitter was placed at the northern end of the path to allow reception of transmissions at the Radio Research Station, Slough.

(2.2) Tests to Determine Propagation Characteristics

The first American results reported had been for tests on

approximately 50 Mc/s,[6] a frequency in the British television broadcasting Band I. It was decided to carry out the initial tests on a higher frequency to avoid possible interference to television, and an allocation at 89·05 Mc/s was obtained. Transmissions started using a carrier power of about half a kilowatt, which was subsequently increased to 2 kW. However, the median level of the received signal was at or below the limit of sensitivity of the receiving equipment most of the time.

Since theoretical considerations suggested that more effective transmission would be obtained at lower frequencies, tests were restarted in November, 1952, using a frequency of 41·1 Mc/s; more sensitive receivers and a somewhat higher carrier power (4 kW) were employed. Tests were made at intervals of 2–3 months over a period of 1½ years, each test lasting 2–4 weeks and being more or less continuous. Most of the propagation investigations were carried out at 41·1 Mc/s, but in order to obtain data on the frequency dependence of the propagation characteristics, additional transmissions on 27·27 and 89·05 Mc/s were provided during two series of tests.* Various aerial systems were provided at Shetland, Jersey and Slough, and a number of investigations were made into the effective gain of directional aerials for this type of propagation.

In October, 1953, measurements in a ship, to provide data on the range and lateral spread of the transmissions, were carried out in collaboration with the Admiralty.

(2.3) Tests on Possible Applications to Communication

A number of telegraphy and telephony tests were made between Shetland and Jersey to determine the suitability of ionospheric-scatter propagation for commercial point-to-point services. The telegraphy tests were confined to frequency-shift operation, but included an investigation of a system of diversity operation which proved to be very satisfactory. For telephony it was considered that phase modulation of relatively low deviation, or single-sideband amplitude modulation, were likely to be the most promising forms of operation, and both were investigated. The telegraphy tests and most of the telephony tests were carried out at 41 Mc/s, but a few single-sideband telephony tests were also made at 27 Mc/s. The received signal levels at 89 Mc/s were generally too low to permit satisfactory telegraphy or telephony tests at that frequency.

* For brevity, these transmission frequencies will subsequently be referred to as 27, 41 and 89 Mc/s.

(3) SITES AND EQUIPMENT

(3.1) Transmitting Station

The transmitting station was located at Ward Hill, near Sumburgh Head, Shetland Isles (Fig. 1). The ground at this site is substantially plane and slopes gradually downwards in the direction of transmission at an angle of about 1·5°; the aerials were about 300 ft above sea level. Details of the main

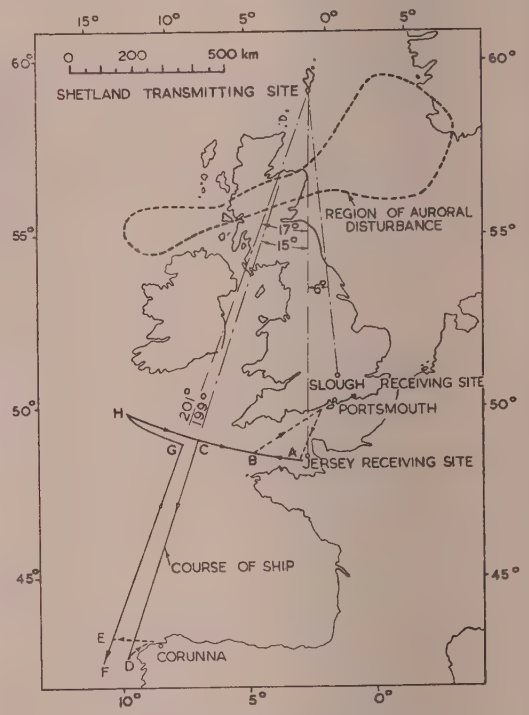


Fig. 1.—Transmission paths and course of ship.

Table 1

TRANSMITTING AND RECEIVING AERIAL DATA

Number	Location	Type	Frequency	Bearing	Plane-wave gain (rel. to λ/2 dipole)	Beam elevation	Beamwidth 3 dB below maximum	
							Elevation	Azimuth
			Mc/s	deg. E of N	dB	deg.	deg.	deg.
1	Shetland	Rhombic 2-stack 14λ side	89	183 (Jersey)	20	5	5·5	7
2	Shetland	Rhombic 2-stack 12λ side	41	183 (Jersey)	19	6	6·5	8
3	Shetland	Rhombic 2-stack 7·5λ side	27	183 (Jersey)	15	10	9	10
4	Shetland	Rhombic 2-stack 10λ side	89	177 (Slough)	17·5	8	7	10
5	Shetland	Rhombic 2-stack 12λ side	41	177 (Slough)	18	8	8	9
6	Shetland	Rhombic 2-stack 12λ side	41	201 (Ship)	19	5	5	8
7	Jersey	Rhombic 2-stack 14λ side	89	3 (Shetland)	20	5	5·5	7
8	Jersey	Rhombic 2-stack 12λ side	41	3 (Shetland)	19	6	6·5	8
9	Jersey	Rhombic 2-singles 12λ side	41	0·5 and 5·5 (Shetland)	17	6	7	8
10	Slough	Rhombic 2-stack 14λ side	89	357 (Shetland)	18	8	7	10
11	Slough	Yagi 4-stack, 3-element	41	357 (Shetland)	12	8	7	50
12	Slough	Yagi single 3-element	27	357 (Shetland)	6	8	8	50
13	H.M.S. *Fleetwood*	Yagi single 3-element	41	Steerable	6	—	—	50

Note: At each fixed station horizontal λ/2 dipoles were also available for use at 27 and 41 Mc/s; vertical λ/2 dipoles were available at 41 Mc/s.

aerials used are given in Table 1; the gains and beam elevations shown are calculated from the aerial dimensions and the height above local ground.

The transmitter used for all the 41 Mc/s tests was capable of delivering an unmodulated carrier power of 4 kW, and facilities were provided for the following types of modulation:

(*a*) Frequency-shift telegraph (f.s.t.) using shifts of either 200 or 800 c/s.
(*b*) Phase modulation (p.m.) by speech or by tones within the frequency range 100–5 000 c/s, at deviations up to ± 2 rad.
(*c*) Single-sideband (s.s.b.) suppressed-carrier amplitude modulation, the upper sideband being radiated and the peak sideband power being limited to about 3 kW.

Although the early 89 Mc/s tests were carried out with lower power, all the measurements at that frequency discussed in the paper were made with an unmodulated carrier power of 4 kW. The transmitter used at 27 Mc/s was designed for independent-sideband suppressed-carrier amplitude modulation. The peak sideband power was about 1 kW, and when used for propagation measurements the same carrier power was employed.

Telegraphy signals were obtained from a standard 50-baud teleprinter or from a tape sender; the latter was normally used with a loop of tape to provide a repetitive test message. Tone modulation was provided from an audio-frequency oscillator, and speech signals were derived from a normal telephone or from tape recordings.

(3.2) Receiving Stations

(3.2.1) Jersey.

The Jersey receiving station was established at Les Landes, near Grosnez Point in the north-west corner of the island. The ground upon which the aerials were erected is almost plane, and slopes downwards at about 1° to the horizontal in the direction of the transmitting station for a short distance and then ends abruptly in cliffs. The aerials were 350 ft above sea level, and details of the types used are given in Table 1.

At Jersey double-frequency-change receivers with a second intermediate frequency of 100 kc/s were used, with an electro-mechanical system of automatic frequency control. For propagation investigations the half-power bandwidth of the second i.f. amplifier was about 70 c/s, thus making possible the accurate measurement of signals down to at least 30 dB below 1 μV, and the indication of signal levels down to about 40 dB below 1 μV across a 75-ohm input. The receivers operated paper-chart recording milliammeters; an automatic level analyser was also available. The latter integrated on self-starting clocks the total time for which the signal had exceeded six predetermined levels, and arrangements were provided for photographing the clock faces automatically every hour.

The f.s.t. receiving equipment was arranged for single-aerial dual-diversity reception with automatic selection of the best output of a pair of aerials, the selection being carried out by a fast-operating level-controlled electronic switch. Frequency shifts of 200 and 800 c/s were used, the i.f. bandwidths being 300 and 900 c/s respectively. It was found to be essential for optimum operation that the bandwidth be restricted prior to the level-operated electronic switch. The output of the f.s.t. receiver was applied to a normal teleprinter (type No. 7B).

For the telephony tests a half-power i.f. bandwidth of 5 kc/s was used for phase-modulation reception at 41 Mc/s. In the single-sideband tests demodulation by local carrier was always employed, and the effective bandwidth was restricted at audio frequencies by a 2·8 kc/s low-pass filter.

(3.2.2) Slough.

The Slough receiving station was at the D.S.I.R. Radio Research Station, where the local ground is substantially plane and horizontal; again, aerial details are given in Table 1. Narrow-band receivers of a triple-frequency-change type were used, the final intermediate frequency being about 1 kc/s, with an audio-frequency filter of 80 c/s half-power bandwidth before the detector. These receivers also operated paper-chart recording milliammeters and used electro-mechanical automatic frequency control.

(3.2.3) H.M.S. *Fleetwood.*

For the range and beam-spread tests a mobile receiving station was provided in H.M.S. *Fleetwood.* Aerial details are given in Table 1; this aerial, which was rotatable, was erected on the top of the foremast about 70 ft above the sea. The recording receiver and automatic level analyser installed were identical in design and performance to those used at Jersey.

(4) GENERAL CHARACTERISTICS OF RECEIVED SIGNAL AT 41·1 Mc/s

A considerable proportion of the total period of the tests was devoted to the examination of the nature of the received signal when an unmodulated 41 Mc/s carrier was being transmitted, and to recording the received signal level for subsequent statistical analysis. A time-constant of about 1 sec was used for the chart recordings, and the resultant smoothing, or suppression of very rapid upward and downward variations, must be allowed for in interpreting the chart data. Throughout the paper signal levels are quoted in decibels relative to 1 μV across the receiver 75-ohm input.

The received signal exhibited a number of unusual features, outstanding among which was its extremely variable amplitude. Large, rapid and erratic fluctuations were almost always present. It is convenient in describing the results to separate those phenomena which were observable on a more or less regular basis and those which occurred relatively infrequently.

The most important regular characteristics may be summarized as follows:

(*a*) For most of the time the signal amplitude varied rapidly over some tens of decibels in a manner resembling random noise of a few cycles per second bandwidth. The amplitude variations were accompanied by random frequency fluctuations of small extent. The median level of this "background signal," which was always present, showed large diurnal and seasonal variations.
(*b*) Frequently the signal amplitude rose suddenly to a much higher value and then dropped more slowly to the background level. The increase in amplitude was up to 40 dB or more, while the duration of the enhancement (or "burst") was generally between about one second and half a minute. These high-level burst-type signals were often accompanied by Doppler frequency shifts of short duration.
(*c*) In addition to the low levels of the instantaneous signal amplitude to be expected from the nature of a randomly varying signal and which occur for a small percentage of the time, short very deep fades lasting a very small fraction of a second were frequently observed.

Less frequently observed phenomena included the following:

(*d*) During daylight hours, very-high-level signals 60–100 dB above the background signal were sometimes received for periods of from several minutes to an hour or more.
(*e*) On one occasion moderately-high-level signals were observed in which the fluctuations of phase and amplitude were much more rapid than normal and of a very erratic character.

(4.1) Instantaneous Amplitude and Phase Fluctuations

What may for convenience be termed the normal amplitude fluctuations of the 41 Mc/s received signal, as recorded with a time-constant of about 1 sec, have been observed to occur over the range of +40 dB to about −35 dB relative to 1 μV; this range includes the slow variations of median value dealt with in the next two Sections. Neglecting the slow variations, the range of fluctuation observed was generally about ± 10 dB and rarely

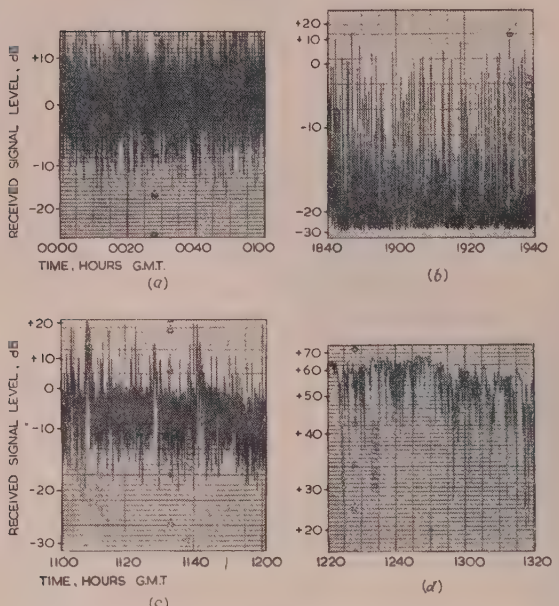


Fig. 2.—Samples of chart records at 41 Mc/s.

(*a*) High-level signal (24.6.53).
(*b*) Low-level signal (20.1.53).
(*c*) Examples of burst type signals (6.5.53).
(*d*) Sporadic-E signal (26.6.53).

more than ±20 dB. The type of chart record obtained, using a chart speed of 3 in/h, is shown in Figs. 2(*a*) and 2(*b*). Level variations occurred at a rate of about 10 dB/sec for most of the time, with level fluctuations of more than about ±6 dB occurring at rates of about 30 per minute.

The type of signal-level fluctuation without smoothing may be seen in more detail in Fig. 3. The traces shown were obtained by photographing the 100 kc/s i.f. signal, using an oscillograph and a motor-driven 35 mm film camera.

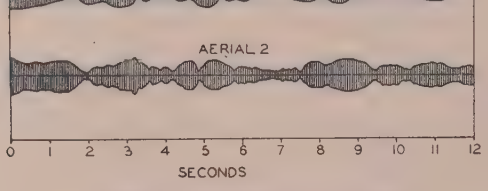


Fig. 3.—Envelopes of i.f. signals.

Showing simultaneous reception on two aerials offset by ±2·5° relative to great-circle path (41 Mc/s).

The time distributions of the received signal levels for a few days during six of the main transmission periods are shown in Fig. 4. These were obtained using the 41 Mc/s rhombic aerial and automatic level analyser at Jersey. Considerable variations in the type of distribution will be noted, but the general tendency is for the distribution to be log-normal, except perhaps for the low-level signals such as are shown by the curve for March, 1953. It may be seen from the curves that the median value ranged from −17 dB in March to +1 dB in June, the overall median being about −9 dB relative to 1 μV. The occasional long-duration high-level signals discussed in Section 4.5 have been excluded from the data used in preparing these distribution curves.

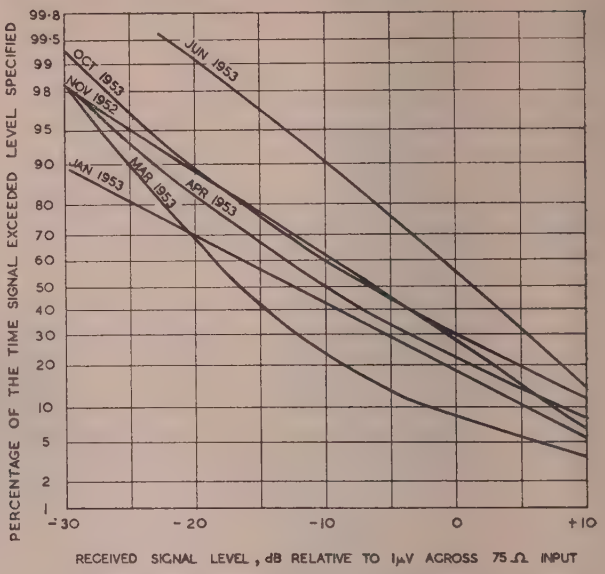


Fig. 4.—Time distributions of received signal levels (41 Mc/s).

Observations have shown that the random variations of amplitude are accompanied by random phase fluctuations of the received carrier. The rate of variation of phase is such that the corresponding frequency deviation rarely exceeds ±10 c/s when normal scatter propagation conditions prevail.

Most of the characteristics of the normal signal during the hours of darkness can be accounted for adequately on the basis that the received signal is the resultant of components scattered from trails of ionization due to numerous meteors. The vast majority of these trails are produced by meteors consisting of extremely fine dust, so their lengths are relatively short, the individual amounts of ionization are small and they are of comparatively low persistence; but their numbers are vast. Occasional larger meteors give rise to much more intense ionization, and the lengths and durations of their trails will be much greater. The magnitude of a received component is, of course, dependent also on the orientation of the particular trail involved, the polarization employed and the location of the trail in space relative to the polar diagrams of transmitting and receiving aerials.[8,15-18] If the transmitted c.w. carrier is taken as a stationary reference vector, components reaching the receiving aerial may be regarded as small vectors varying rapidly in phase and amplitude, and the received signal will be their vector sum, i.e. a vector of almost random magnitude and phase relative to the reference vector. Since individual trails are created at finite rates and take an appreciable fraction of a second or longer to die away, changes in magnitude and phase of this resultant signal, although rapid, occur at finite rates.

It is thus believed that the night-time median signal level is dependent primarily on the number, size, distribution and predominant direction of meteors arriving at that time in the parts of the lower ionosphere illuminated by the main and subsidiary lobes of the transmitting and receiving aerial systems.

During daylight the ionization of the E-region increases considerably and the significance of turbulent scattering increases in relation to that from meteor trails.

(4.2) Diurnal Variations

An estimate of the median value of the background signal received at Slough and Jersey has been made for each hour of

the transmissions by visual inspection of the recorder charts. The variation of this median value with time of day is shown in Fig. 5 for two of the monthly transmission periods. The

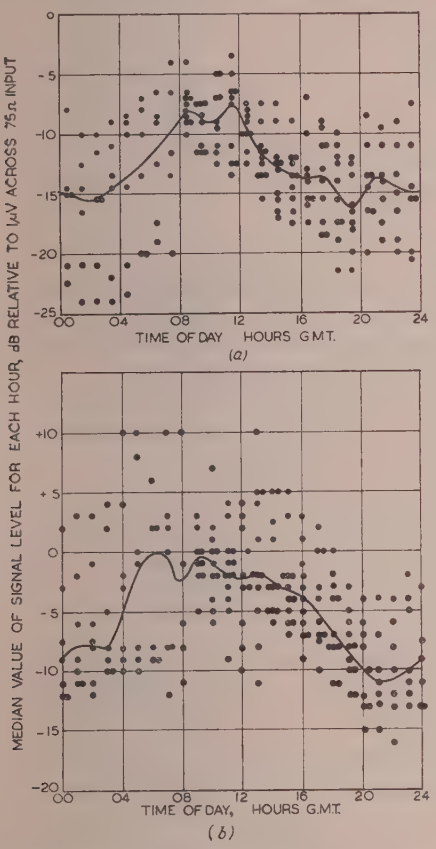


Fig. 5.—Diurnal variation of median values of signal level for each hour (41 Mc/s).

(*a*) Received at Slough (April 27 to May 8, 1953).
(*b*) Received at Jersey (June 18 to July 7, 1953).

maximum day-to-day scatter of the median signal level at any particular hour of the day is from 10 to 20 dB over a monthly period, which indicates the variable nature of this form of propagation. The full-line curves show how the median value for each test period varied throughout the day.

Curves for reception at Jersey and Slough during the period November, 1952, to March, 1954, similar to the full-line curves in Fig. 5, are shown in Fig. 6. Certain regular features are apparent from this diagram. The diurnal maximum is during daylight and, when clearly defined, is between the hours of 1000 and 1300*; this maximum is no doubt due to the predominance of turbulent scattering when the average electron density in the E-region is at its greatest. The duration of the maximum varies with the period of daylight and is generally longer in summer than in winter; it is to be noted, however, that the increased duration in summer occurs mainly in the forenoon rather than in the afternoon. A diurnal minimum usually occurs at about 2000 hours, irrespective of the period of daylight; this corresponds to the time of minimum meteoric ionization. A comparison of the diurnal curves for reception at Jersey and Slough shows that there is, in general, a greater variation between maximum and minimum signal levels for the longer path, i.e. the path to Jersey.

* All times are G.M.T

(4.3) Seasonal Variations

The median of the hourly median received signal levels at Jersey and Slough has been determined for each test period and plotted in Fig. 7. Although not extensive, these results give an indication of seasonal variation. From the Slough measurements it can be seen that there are minima at the equinoxes and maxima in summer and winter; the same is true for reception at Jersey, except that the autumnal minimum is missing. These variations are generally similar to the seasonal changes in the occurrence of sporadic-E ionization. There is also evidence of correlation with meteoric activity. For example, the lowest median signal levels were observed in March, when it is known that both the day and night rates of occurrence of meteors are lower than at other times of the year; the highest median signals were obtained during daytime in June, coincident with the annual maximum incidence of meteors.

In addition to the seasonal variations there appears to be a general downward trend over the whole length of time of the measurements. This secular variation might be associated with the sunspot cycle which passed through its minimum in 1954.

(4.4) Short-Duration Abnormal Fluctuations

Numerous short-duration enhancements in the signal level have been observed throughout the tests. A specimen chart recording showing several of these burst-type signals is given in Fig. 2(*c*). The rise in level was of the order of tens of decibels and was generally completed within a second. The decay was slower, and for the larger bursts it was normally oscillatory. Durations of as much as 10 sec were fairly common, and some of over a minute were occasionally observed; durations of 1–30 sec may be taken as representative. No precise lower limit for amplitude or duration can be given, as the smaller and shorter bursts merge imperceptibly into the random background fluctuations already described. Large bursts were always of appreciable duration, but the smaller-amplitude bursts were not invariably of short duration.

The rise in level was generally preceded, or accompanied, by a whistle of descending pitch and occasionally of ascending pitch, but many enhancements were observed without any audible whistle. At times the note was heard to descend in pitch and then rise again. There is little doubt that some of these whistles owe their origin to the beat between the background scatter signal and a signal scattered from a moving head of meteor ionization, which undergoes a Doppler frequency-shift. Other whistles can arise from the amplitude modulation of the received signal caused by the changing multi-lobed reradiation pattern associated with a column of ionization of varying length; alternatively this phenomenon may be regarded as due to the growth of a Fresnel diffraction pattern.*

Bursts accompanied by an audible whistle are undoubtedly caused by the arrival of a meteor. It is possible that another mechanism is required to account for those bursts which are not accompanied by an audible whistle, but the authors consider that the meteor-trail theory is adequate. The absence of a whistle can almost certainly be accounted for in terms of the following factors:

(*a*) The geometry of the transmission path and the meteor trail being such that the rate of change of path length is small.
(*b*) The directivity of the transmitting and receiving aerials.
(*c*) The low-frequency cut-off of the receiver and loudspeaker in use.

The rate of rise of burst-type signals was normally extremely rapid, and rates as high as 100–200 dB/sec have been measured. The maximum amplitude could be 60–80 dB or more above the

* See Chapter VIII of Reference 19.

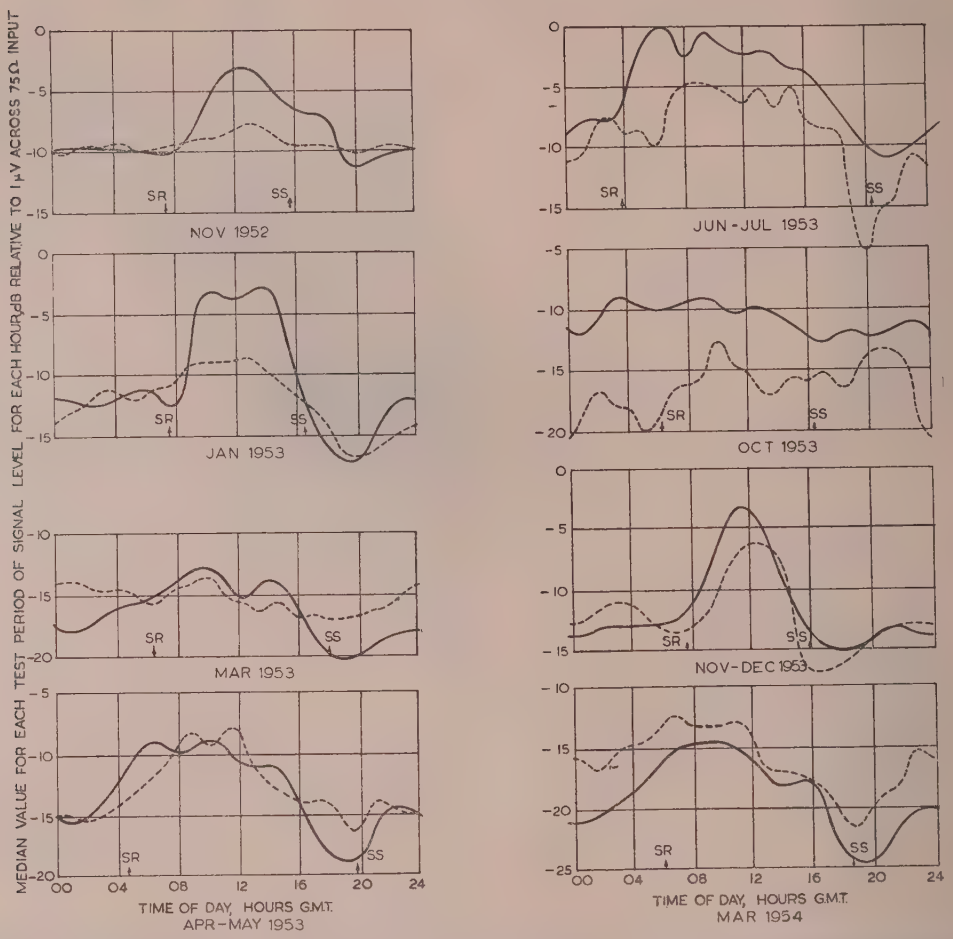


Fig. 6.—Median value for each test period of signal level for each hour during the period November, 1952, to March, 1954 (41 Mc/s).

——— Received at Jersey.
– – – – – Received at Slough.
S.R. Time of sunrise at mid-path.
S.S. Time of sunset at mid-path.

prevailing background level for the larger bursts, although rises of 20–40 dB were much more frequent.

Provided that the receiving equipment had a rapid enough response, very-short-duration deep fades were a conspicuous characteristic of the received signal. These were clearly distinguishable from the downward fluctuations to be expected in a truly random signal, and as their durations were of the order of milliseconds they do not appear in the chart records. They are of very great importance in telegraphy circuits relying on this type of propagation.

These fades were randomly distributed in time, and the mean interval between fades ranged from some five or ten seconds to several minutes. More frequent fading generally coincided with higher-than-average numbers of burst-type signals and meteor whistles, and the shape of the signal envelope of these short fades showed the usual characteristics of wave-interference fading. They were frequently directly associated with particular bursts, an extremely fast "drop out" (of the order of a few milliseconds) preceding some enhancements, and one or more fades (of the order of tens of milliseconds) occurring during the decay of some of the smaller and relatively short-duration enhancements.

There is little doubt that the short-duration fades are due to momentary cancellation of the prevailing background signal by

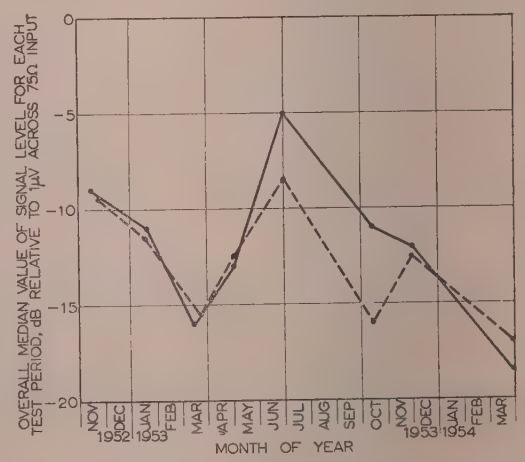


Fig. 7.—Variation, with time of year, of overall median value of signal level for each test period, November, 1952, to March, 1954 (41 Mc/s).

——— Received at Jersey.
– – – – – Received at Slough.

the signal received from one of the larger meteors. If the transmitted carrier is taken as the reference vector the background signal can be regarded as a vector varying in amplitude and phase in more or less random manner but at a moderate rate; a meteor arriving in a suitable direction can provide a signal which is then represented by a vector rotating at quite high speed (e.g. at a few hundred cycles per second) and growing rapidly in amplitude, and it is probable that on one or more of its rotations it will have about the right magnitude partially or completely to cancel the background vector as it passes through phase opposition, and one or more very short breaks in reception will occur.

(4.5) Long-Duration High-Level Signals

On many occasions during the tests very strong signals lasting for a few minutes or more were received, while some lasted for periods up to 3 hours. High-level signals of this type are shown in Fig. 2(*d*); they are believed to be due to sporadic-E ionization and are distinguished from meteor bursts by the relatively slow initial rise in strength. On one occasion the signal level rose to within 10 dB of the theoretical free-space value, which is about 100 dB above the normal level of the scattered signal. The signal-level variations showed deep fades of at least 30–40 dB and rounded maxima. All these characteristics indicate that propagation was taking place by a mechanism approximating to specular reflection such as might occur from a sporadic-E cloud or clouds suitably situated along the transmission path. The interference type of fading shown in Fig. 2(*d*) suggests the existence of more than one propagation path. Back-scatter measurements, with a pulsed transmitter at 17 Mc/s, made at Slough during one of these periods of strong signal, showed a sporadic-E cloud some 250 km in extent and 500 km due north of Slough, i.e. about the mid-point of the path from Shetland to Slough.

Periods of high signal exceeding a few minutes' duration occurred most frequently during daylight and were more numerous during summer and winter than in spring and autumn, which underlines their close association with the occurrence of sporadic-E clouds as shown by vertical-incidence measurements at similar latitudes.

(4.6) Auroral Effects

On 15th October, 1953, between about 1630 and 1830 hours, quite abnormal propagation conditions were observed at both the fixed receiving stations and also in H.M.S. *Fleetwood*, which was then carrying out the range tests described in Section 7. The signal-level variations became more rapid than usual until the signal was fluctuating violently, the phenomenon resembling the flutter-fading well known in short-wave communication.

The chart record obtained at Slough is shown in Fig. 8; similar records were obtained at Jersey and in H.M.S. *Fleetwood*,

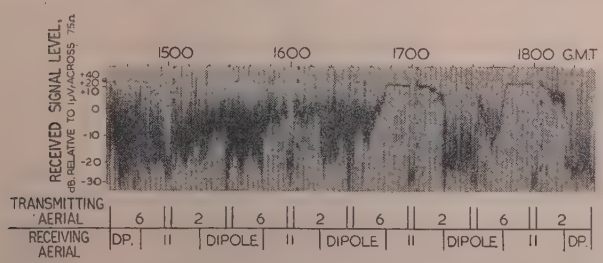


Fig. 8.—Chart record of received signal (41 Mc/s) at Slough showing effects of an auroral disturbance on 15th October, 1953.

The various transmitting and receiving aerials in use during the period of observation are indicated below the record in accordance with the numbers given in Table 1.

which at the time was near point C in Fig. 1. The level fluctuations were large, but occurred so rapidly that they were considerably smoothed by the mechanical and electrical time-constants of the recording equipment, and the thinning of the trace which will be observed could well be mistaken for abnormally small fluctuations. The general signal level reached maxima at about 1650 and 1750 hours and then fell gradually to a normal value. Listening tests showed that there was considerable random amplitude and frequency modulation of the received carrier, and observations with a spectrum analyser at Slough showed that sidebands extended up to about ± 200 c/s. While the increase of level and its duration at the more distant points, Jersey and H.M.S. *Fleetwood*, was smaller than that at Slough, it is evident that the effects of the disturbance were widespread. It was not practicable to carry out telegraphy or telephony tests at short notice while the phenomenon was still in progress, but there is little doubt that it would have seriously degraded or completely interrupted any service.

This abnormal propagation occurred at the start of severe ionospheric and magnetic storms, and it was subsequently established[20] that at about 1800 hours an auroral disturbance was located across the propagation path but some distance north of the mid-point of the path. Data on the approximate shape and location of the auroral disturbance at 1800 hours are shown in Fig. 1. It may be significant that a solar flare and a sudden ionospheric disturbance occurred at about 1000 hours on the previous day, i.e. some 30 hours earlier.

(4.7) Polarization of Received Signal

A few brief investigations were made of the polarization of the received signal and the relative merits of using horizontal or vertical polarization at the transmitting station. The results indicated that the transmission loss, P_r/P_t, was smaller when using horizontal polarization, and that horizontal polarization is substantially preserved during propagation. With nominally vertically polarized transmissions the horizontally polarized component of the received signal was large, and during daylight hours it sometimes predominated over the vertical component.

The advantages of horizontal polarization were quite substantial in tests to Jersey; similar results were obtained at Slough, but the predominance of horizontally polarized components in the received signal when transmitting with vertical polarization was not quite so marked. Results varied considerably with time of day and from day to day, and many more tests over a long period would be required before a reliable statistical result could be established.

(4.8) Wave-Angle Diversity

In view of the random nature of the signals it was considered that receiving aerials aimed at different parts of the scattering region should give signals in which the fine structure of the fluctuations would be more or less uncorrelated, although the median levels would be expected to be approximately the same. Diversity reception was therefore considered to be possible as a means of avoiding many of the short periods of very-low-level signal arising from appreciable downward fluctuations from the median value.

For the first tests at Jersey, two rhombic aerials were used which differed in beam width and main-lobe elevation, but both were aligned along the great-circle direction of the transmitting station. In February, 1953, these were replaced by two rhombic aerials having approximately identical beam widths and main-lobe elevations, but directed $+2{\cdot}5°$ and $-2{\cdot}5°$ relative to the bearing of the transmitting station. This pair of aerials was used for all subsequent wave-angle diversity tests. The effectiveness

of the diversity operation was investigated by observing the envelopes of the i.f. signals in two receivers connected to the aerials under examination; a typical record showing the lack of correlation between the fine structure of the signals is given in Fig. 3. Results for aerials differing in elevation instead of in bearing were much the same.

Recent theoretical work by Eshleman and Manning[10] has shown that for scattering from meteor trails the received signal level should be a maximum when the transmitting and receiving aerials are directed a few degrees to one side or other of the great-circle path; the advantage to be gained by so doing is likely to be appreciable when highly directional aerials are used. With this in mind it would seem that the diversity arrangement in which similar aerials are directed to one side or other of the great-circle path is to be preferred.

Wave-angle diversity reception proved to be a particularly effective method of avoiding most of the short-duration cancellation fades discussed in Section 4.4. Since the background components from the two aerials differ at most times in magnitude and phase, the probability of simultaneous cancellation of both by a spiralling vector due to a meteor is quite small, and the time displacement of cancellations arising on the two aerials is generally sufficient for effective diversity operation, provided that a fast-operating switch is used.

(4.9) Noise Levels

The effective input-circuit noise level due to sources within the 41 Mc/s receivers was about −43 dB relative to 1 μV across 75 ohms; this very low level was achieved by restricting the bandwidth to 70 c/s and using a low-noise input stage. At Jersey the increase in the noise on connecting the aerial was found to vary at different times, but normally lay within the range 2–4 dB. The mean figure was about 3 dB, and the difference between dipole and rhombic aerials was usually under 1 dB. The lowest recordable level at Jersey was thus limited by the ambient noise level of about −40 dB relative to 1 μV. At Slough the prevailing level of local noise was several decibels above that at Jersey, and recordings could not be made to such a low level of signal.

(5) VARIATION OF SIGNAL CHARACTERISTICS WITH FREQUENCY

During November and December, 1953, and March, 1954, transmissions were made at frequencies of 27 Mc/s and 89 Mc/s, in addition to those at a frequency of 41 Mc/s. The same path was used for the three transmissions, so that a comparison of the received signal characteristics could be made and variations in signal intensity with frequency could be studied. The 27 Mc/s transmitter was operated continuously and the 41 and 89 Mc/s transmitters in alternate hours.

(5.1) Comparison of General Characteristics at 27·27, 41·1 and 89·05 Mc/s

The nature of the short-term variations on the three frequencies was very similar, although little correlation was observed in the instantaneous fluctuations of the background signal. A comparison of the amplitudes of these fluctuations shows that they are substantially the same at the three frequencies and in particular that there is a reduction in the amplitude of fluctuation during daylight. A comparison of the diurnal variation of the hourly median signal level was made by inspection of samples of the chart recordings. It was found that the shape of the diurnal curve was very similar for the three frequencies, with maxima and minima occurring at about the same time of day, and that the ranges of variation were approximately equal.

A comparison of the abnormal fluctuations at the three frequencies showed that the rate of occurrence of bursts of signal appeared to decrease with increase in frequency. Large meteor bursts, however, could be identified at each frequency as being due to the same meteor, the tendency being for a sharp rise to occur at 89 Mc/s followed by a fall lasting up to half a minute, while at 27 Mc/s the rise was somewhat slower and the fall lasted up to 3 min. On the few occasions when sporadic-E propagation was observed there was a more continuous high-level signal with fewer fades at 27 Mc/s than at 41 Mc/s. None of these occasions coincided with the operation of the 89 Mc/s transmitter.

(5.2) Variation of Transmission Loss with Frequency

In order to determine the exponent s of the frequency term in expression (1), which indicates how the theoretical transmission loss varies with frequency, observations have been made of the signal level at the various frequencies. In the comparisons between propagation at 89 and 41 Mc/s, transmitting rhombic aerials Nos. 1 and 2 were used at Shetland in conjunction with rhombic receiving aerials Nos. 7 and 8 (identical with Nos. 1 and 2) at Jersey, and with aerials Nos. 10 and 11 (rhombic and Yagi array respectively) at Slough. In the comparisons between propagation at 41 and 27 Mc/s, $\lambda/2$ dipoles were used at Shetland and Jersey; the same applied at Slough with the exception that Yagi aerial No. 12 was used for reception at 27 Mc/s. The 41 and 27 Mc/s median signal levels could be compared for the same interval of time, but the 89 and 41 Mc/s median levels could be compared only in adjacent intervals, since both transmitters could not be operated together. Where necessary, corrections for differences in aerial gain, transmitter power and feeder losses were made when comparing median levels.

During November and December, 1953, about 100 median measurements were made both day and night for each frequency comparison over a period of 12 days. The distribution of the exponent with time for the Slough and Jersey observations, after allowance for the factors mentioned above, is shown in Table 2.

Table 2

DISTRIBUTION WITH TIME OF FREQUENCY EXPONENT OF TRANSMISSION LOSS

For systems with aerials scaled according to the wavelength.

Percentage of time value of exponent is equalled or exceeded	Exponent of frequency, s			
	89 : 41 Mc/s		41 : 27 Mc/s	
	Jersey	Slough	Jersey	Slough
%				
5	7·5	8	4·5	4·5
10	7	7·5	4	4
50	5·5	6	2	2
90	4·5	4·5	+0·5	0
95	4	4	0	−1
Number of observations	120	100	110	100

It is apparent that the exponent is not constant but depends on the frequencies being compared. Thus the concept of an exponent of frequency in the transmission-loss expression is of limited value, and it is more useful to relate the median received power to the frequency on the basis of aerials scaled according to the wavelength (i.e. of equal theoretical gains) and for a constant transmitted power. The observed variation of trans-

mission loss with frequency is shown in Fig. 9; the variation was the same at Jersey and Slough, which suggests that the curve in Fig. 9 is applicable for ranges of at least 900–1 200 km.

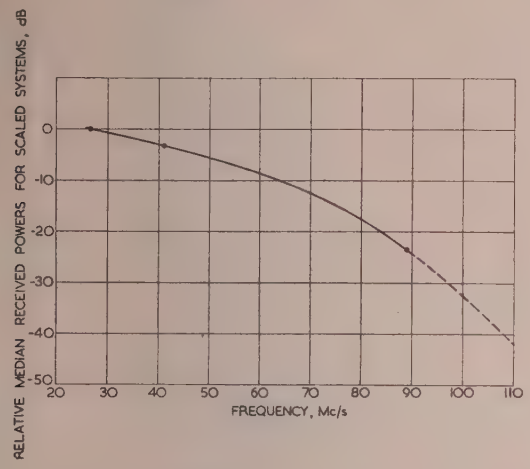


Fig. 9.—Variation of transmission loss with frequency for system with aerials having dimensions scaled according to the wavelength.

The curve shows a rapid decrease in received power as the frequency is increased and indicates that this form of propagation is most favourable in the range approximately 25–60 Mc/s. The distributions with time of the relative transmission losses of systems using scaled aerials at frequencies of 27 and 41 Mc/s, and 41 and 89 Mc/s, are shown in Fig. 10; the data shown are those for reception at Jersey, and are based on a comparison of the hourly median values.

The diurnal variation of the frequency dependence of the received power has been investigated, and it was found that at both receiving stations and for both frequency comparisons there was, in the day-time as compared with the night-time, a slight increase in the advantage to be obtained by using the lower frequency.

Since the signal levels themselves have generally been found to be higher during the day-time, the correlation of the advantage of the lower frequencies with signal level has been examined; it was found that in general slightly greater advantages occurred with the higher signal levels.

The conclusions reached from the November–December, 1953, measurements were substantially confirmed by a second series made during nine days in March, 1954. However, in considering the above results it should be borne in mind that the 27 to 41 Mc/s comparisons were made between dipole aerials, while the 41 and 89 Mc/s comparisons were made between rhombic aerials.

(5.3) Observed Dependence of Transmission Loss on Frequency in Relation to Theory

The interpretation of the results on the frequency exponent, s, of transmission loss, given in Table 2, in terms of the theories outlined in Section 13.1, presents considerable difficulties. Although the range of variation of the exponent obtained in the comparison of either pair of frequencies might be attributed to changes in the scale of turbulence, apart from the low values of 0 and -1 recorded for the comparison 41 : 27 Mc/s, the same scales of turbulence will not satisfy both pairs of frequencies. (The scale of turbulence may be thought of as roughly representing the size of the turbulent eddies, and must be approximately the same as the wavelength in order to produce variations

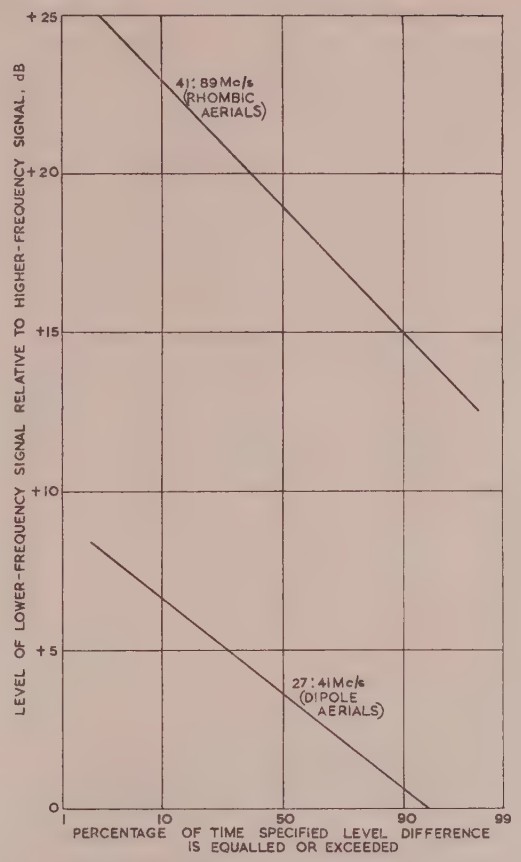


Fig. 10.—Distribution with time of relative transmission loss of systems with scaled aerials, operating at 27 : 41 Mc/s and 41 : 89 Mc/s.

in s.) It might be suggested that variations in s could also occur as the constitution of the received signal varied from being predominantly due to scattering from meteor trails to being primarily the result of scattering caused by turbulent fluctuations of electron density; but in this case it is difficult to see why the exponent in the 41 : 27 Mc/s comparison is so much less than in the 89 : 41 Mc/s comparison. A further complication is that, since the scale of turbulence must be of the order of the wavelength in the frequency range concerned, the power received over a given link will be materially affected by changes in the radiation patterns of the scattering centres, quite apart from the effect of the frequency factor itself.

The results obtained in the comparison of the 89 and 41 Mc/s transmissions could be accounted for—qualitatively, at least—on the assumption that the scattering is due to both turbulence and to meteoric ionization, the relative contributions of these two agencies being variable, and on the further hypothesis that the spectrum of turbulence contains scales as small as 10 m for an appreciable fraction of the time, as has recently been shown by Maxwell.[21] If this explanation is correct it appears that the power received at 27 Mc/s is for some reason lower than would have been expected; and it seems worth while considering whether absorption in the ionosphere could account for the discrepancy, since such absorption must become increasingly significant as the frequency is reduced.

There is little direct evidence concerning the precise magnitude of the absorption to be expected at very high frequencies for the mechanisms of transmission under consideration, and an attempt

has therefore been made to assess the importance of such absorption from a knowledge of collision frequencies and electron densities in the part of the ionosphere involved. It seems probable from the investigation of aerial performance described in Section 6 that scattering takes place over a considerable range of height in the ionosphere, say from about 80 up to about 120 km, and is not confined to a thin stratum of only 5 km or so thickness, as is assumed by Bailey and his colleagues.[6] Scattering from meteor trails would certainly be expected to occur over a considerable range of height, and the same must also be true for scattering due to turbulence, since, so far as the authors are aware, there is no evidence that winds in the E-region are confined to a relatively thin layer.

Recent day-time rocket investigations in America[22,23] show that the ionization density rises steeply with height from a very small value near to a height of 80 km, and that above about 100 km the ionosphere is a heavily ionized continuum. The idea of separate layers must be replaced by the more appropriate concept of adjacent E_1, E_2, F_1 and F_2 regions blending gradually into one another. In the day-time the E_1 and F_2 layers have large electron density gradients on their lower sides. From these investigations it appears that the electron density is of the order of 5×10^{10} per cubic metre at 90–95 km height; and over the range from 100–130 km lies between $1{\cdot}0$ and $1{\cdot}7 \times 10^{11}$ per cubic metre (for a typical sounding), there being a local maximum at the height of the E_1 layer. The collision frequency, of course, varies greatly over this range of height, being of the order of 10^6 per second at 90 km, 5×10^5 per second at 100 km and 10^5 per second at 120 km; most of the absorption must therefore take place in the lower E-region, despite the high density of ionization which persists throughout the whole region.

Calculation shows that the difference in absorption to be expected at 27 and 41 Mc/s under the conditions described above is of the order of 2 dB for scattering at heights up to, say, 100 km; but over a path involving scattering at 120 km or more the differential absorption might be as much as 4–5 dB. The average effect of absorption over a large scattering volume lying in the range 80–120 or 130 km, therefore, may well be such as to cause a difference in attenuation at 27 and 41 Mc/s in the region of 3–4 dB, the greater absorption occurring at the lower frequency, although it is obviously difficult to arrive at any precise figure for this difference. An effect of the above magnitude is sufficient to reduce the frequency exponent in the comparison of scattered powers at 27 and 41 Mc/s by a value of up to 3, which is practically the whole of the discrepancy between the observed exponent for these two frequencies and that obtained in the comparison between transmissions on 89 and 41 Mc/s. It should be added that the small absorption occurring at 41 Mc/s, in relation to the still smaller absorption occurring at 89 Mc/s, is such as to produce no important change (not more than 1) in the value of s obtained from measurements at the two higher frequencies.

Since values as low as 3 or 4 are obtained at times for the exponent s in expression (1), corresponding to only 1 or 2 for the exponent of f in the expression for scattered power, $\sigma(\theta, \chi)$ (see Section 13.1), it is probable that at such times the major part of the signal is due to turbulent scattering, with the scale of turbulence of the same order of magnitude as the wavelength. For larger scales of turbulence the frequency exponent increases, but from the general character of the received signal it is likely that higher values of s also indicate an increasing contribution to the total signal from radiation scattered by meteor trails.

(5.4) Selective Fading

The extent of frequency-selective fading when propagation is by ionospheric scattering has been investigated. Each sideband channel of the 27 Mc/s transmitter was modulated by a sinusoidal tone, and the relative levels of the received tones were examined after demodulation in the separate sideband channels of the receiver. The output tones were displayed on a double-beam cathode-ray oscillograph after being passed through narrow-band filters to reduce the noise. The receiver gain was controlled by a pilot carrier, thus almost eliminating simultaneous variations in the levels of the sideband frequencies.

Two tests were made using tones of 5 and 1 kc/s, thus placing the radiated frequencies 10 and 2 kc/s apart respectively. It was observed that considerable selective fading was present with the 5 kc/s tone while very little occurred with the 1 kc/s tone. It may be concluded that selective fading is to be expected in propagation by ionospheric scattering; it follows that frequency diversity with a frequency spacing of about 10 kc/s or more might improve the usefulness of this type of propagation for telegraph transmission.

(6) INVESTIGATION OF AERIAL PERFORMANCE

A study has been made of the performance of transmitting and receiving aerials when propagation is by ionospheric scattering. The measurements were made at 41 Mc/s, the transmitting aerials at Shetland being a $\lambda/2$ dipole and a rhombic aerial (No. 2 in Table 1). At Jersey similar aerials were used for reception, but at Slough, where most of this study was carried out, the aerials available were a $\lambda/2$ dipole and a stacked Yagi array (No. 11 in Table 1). By suitably switching aerials at the transmitting station and at Slough, measurements of the median signal level were obtained in adjacent 15-min intervals for all four combinations of transmitting and receiving aerials. The tests were repeated in several monthly transmission periods.

The effective gain of the receiving array at Slough relative to the $\lambda/2$ dipole at the same height was found to vary over the range 0–12 dB. The distribution with time of the measured gain of the receiving array when a rhombic aerial was used at the transmitting end is given in Table 3 for four separate transmission periods and shows little variation with the time of year.

Table 3

MEASURED GAIN OF RECEIVING ARRAY AT SLOUGH

Percentage of time the stated gain was equalled or exceeded	Measured gain of receiving array when transmitting from a rhombic aerial				
	June–July, 1953	Oct., 1953	Nov.–Dec., 1953	Mar., 1954	Overall
%	dB	dB	dB	dB	dB
5	10	11	9½	10½	10½
10	9½	10	9	9½	9½
50	6½	6	6	6	6
90	3½	2	3	2½	2½
95	2	1	2	2	1
Number of observations	200	270	100	80	650

It can be seen that the median measured gain of the receiving array is about 6 dB, and is about one-half of the theoretical plane-wave gain in decibels. Similar measurements at Jersey yielded a median value of 10 dB for an aerial of theoretical gain 17 dB. These results suggest that for this form of propagation the median effective power gain of a highly directional receiving aerial, when a similar aerial is used at the transmitting end, is approximately equal to the square root of the theoretical power gain. It is shown in Section 13.2 that, for highly directional aerials

Such a result is to be expected when the scattering is due to turbulence, the scattering layer is thick enough to embrace the region of overlap of the transmitting and receiving aerial beams, and $G_t \geqslant G_r$. The idealized conditions assumed in the analysis in Section 13.2 are more nearly met for the rhombic-to-rhombic transmissions than for the rhombic-to-Yagi-array transmissions, since the beam width of the array is appreciably greater in azimuth than in elevation; but, even allowing for this, the fact that the full gain of the receiving aerial is seldom realized suggests that for most of the time the scattering layer cannot be thin in relation to the depth of overlap of the aerial beams. Effective gains in excess of the median value of $\sqrt{G_r}$ could be accounted for by a reduction in the thickness of the turbulent layer until it was comparable with, or at times less than, the depth of overlap of the beams. The gain to be expected when scattering from meteor trails is predominant is less certain, since the distributions in time and space of all trails able to contribute to the signal are unknown; but for completely random distributions the effective gain according to Eshleman and Manning[10] may be very low, and this might explain the low values recorded in Table 3.

However, when the $\lambda/2$ dipole was being used for transmitting instead of the rhombic, the median measured gain of the receiving array at Slough was found to be only 4–5 dB. If the discussion in Section 13.2 were directly applicable in this case, where $G_r > G_t$, no gain would be expected from the Yagi array. However, the radiation pattern assumed in the analysis is very different from that of a dipole, and it is certain that, for turbulent scattering at least, relatively little contribution to the received signal can come from regions of the ionosphere illuminated at large angles away from the great-circle path, provided that the scale of turbulence is not very much less than the wavelength, and this is believed to be so. Something should therefore be gained by focusing the receiving-aerial beam on the more important scattering region illuminated by the transmitting dipole near to the great-circle path.

No very marked diurnal variation of the effective gain of the high-gain receiving aerials at either Jersey or Slough was observed, although there was some tendency to a minimum around 2000 hours. The correlation between effective aerial gain and signal level has also been investigated; the hourly median gain of the Slough array during the period June–July, 1953, is shown plotted against the corresponding values of the median signal level in Fig. 11. It can be seen that there is a tendency for the

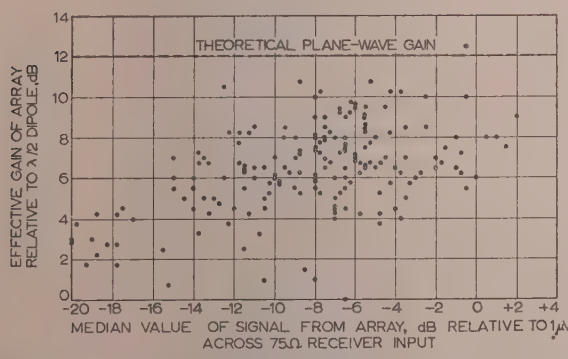


Fig. 11.—Variation of effective gain of Yagi array (No. 11) with median value of signal level as received by the array (41 Mc/s).
Rhombic transmitting aerial No. 2.

effective aerial gain to be less at the lower received signal levels, but this trend may be influenced to some extent by the approach to the noise level.

A set of measurements corresponding to those described above was made for the transmitting rhombic aerial, which had a theoretical plane-wave gain of 18 dB. The median signal level received from this aerial was compared with that received when transmitting from a $\lambda/2$ dipole. Gains of 0–10 dB were observed, and the distribution with time when receiving on the array and on the $\lambda/2$ dipole at Slough are shown in Table 4 for two of the monthly transmission periods.

Table 4

MEASURED GAIN OF TRANSMITTING RHOMBIC AERIAL AT SHETLAND

Percentage of time the stated gain was equalled or exceeded	Measured gain of transmitting rhombic aerial when receiving on a dipole or an array			
	Nov.–Dec., 1953		Mar., 1954	
	Receiving on array	Receiving on dipole	Receiving on array	Receiving on dipole
%	dB	dB	dB	dB
5	8	7½	7½	5½
10	7	6	6½	4½
50	4½	3½	3½	2½
90	2	1½	1	1
95	1	1	1	1
Number of observations	100	100	50	50

The median measured gain is below even one-half the theoretical gain in decibels. It may be significant that similar but rather limited measurements made at Jersey with a rhombic receiving aerial tended to indicate somewhat greater gains at times than shown in Table 4. These facts are not surprising when it is remembered that once $G_t > G_r$ the received power is determined only by G_r; therefore it is not to be expected that a median gain of the transmitting rhombic of 10 dB (the value of a similar receiving rhombic when identical aerials are used) will be realized when reception is by means of the Yagi array. The median gain of the transmitting rhombic relative to a dipole, as measured on the array, cannot be more than that of the receiving array when the transmitting aerial is a dipole, i.e. 4–5 dB, which is what the results in Table 4 show. Once again, if the analysis in Section 13.2 were strictly applicable, when the receiving aerial is a dipole the rhombic transmitting aerial should show no gain over a transmitting dipole. In fact, since all parts of the very large volume defined by the region of overlap of the radiation patterns of the two dipoles are not equally effective from the scattering aspect (which is contrary to the assumption in the analysis), some gain is obtained by concentrating more power in the more effective region near to the great-circle path. The gain of the transmitting rhombic as determined by observations with a receiving dipole is, in general, less than when the Yagi array is used for reception, and this is what would be expected from the theoretical considerations. It therefore appears that in order to achieve the maximum gain from a transmitting aerial it is necessary to use a receiving aerial of similar directivity.

(7) RANGE AND BEAM-SPREAD TESTS

It is of considerable interest, both theoretically and practically, to establish the way in which the received signal varies with path length for propagation by ionospheric scattering, and especially to determine the maximum range at which reliable reception can be achieved. It is also desirable to determine the angular

spread of the scattered beam, i.e. variation of signal strength for receiving sites at the same distance from the transmitter but displaced at various angles from the axis of the beam.

Observations at the Jersey and Slough receiving stations could provide only very limited data on these questions, as the paths did not differ very greatly in length and the angle between them was fixed. A proposal was made by the Admiralty to extend the scope of the measurements by carrying out trials with the receiver in a ship. The Post Office fitted a receiver in H.M.S. *Fleetwood* and provided staff for the measurements of range and beam spread which were carried out in October, 1953. The Admiralty Signal and Radar Establishment collaborated in the planning of the trials, was responsible for the fitting of special aerials in the ship, and provided assistance for the work in H.M.S. *Fleetwood*.

(7.1) Scope of Investigation

The rhombic aerial (No. 6) provided in Shetland for the ship tests was orientated on a bearing of 201° E. of N., along which range tests were to be made. The ship tests took place during the period 6th–17th October, 1953, and provided nine days of continuous 24-hour recording of field strength of the 41 Mc/s transmissions from Shetland. The approximate course followed by H.M.S. *Fleetwood* is shown in Fig. 1, the important parts of the voyage being indicated by solid lines. Over portions AC of the outward voyage and GHCB of the return voyage, the path length was approximately constant and data were obtained on beam spread. Portions CD of the outward voyage and EFEG of the return provided data on the variation of received signal level with distance from the transmitting station. Observations made simultaneously at Jersey were used to correct the ship results for changes of general median level.

During the ship trials, alternate transmissions were made on the Jersey-directed aerial (No. 2) and on the low-angle rhombic on a bearing of 201° E. of N., and both these transmissions were monitored at Slough. The observations at Jersey and Slough as the transmission changed from one aerial to the other also provided information on the variation of the effective spread of the beam for a fixed angle off-beam.

(7.2) Range Tests

(7.2.1) Ship Test Results.

Considerable difficulty was experienced in analysing the range test data. It was expected that the median level measured in the ship, when corrected for diurnal variations as measured in Jersey, would display a reasonably smooth variation with distance, but this was not so. The variation of median signal level with distance (Fig. 12) showed a very large undulatory component, although the median levels measured at the same time at Jersey and Slough were comparatively stable; it seems likely that the multi-lobe structures of the transmitting and receiving aerials due to sea reflections, and possible variations in the height of the scattering region, contributed to this result.

If the undulatory variations are neglected the received signal level decreases relatively slowly with distance up to about 1 900 km; beyond 1 900 km the level decreases rapidly and the signals merge with the receiver noise beyond 2 050 km. The decrease of signal at extreme range corresponds approximately to that to be expected from the vertical-plane radiation diagram of the receiving aerial as determined by its height above sea level, and assuming a scattering region with its lower boundary at about 80 km, due allowance being made for refraction in the troposphere. It follows that, in the application of scatter-signal propagation to very long paths, the design and siting of the aerials is of extreme importance. The height of the transmitting

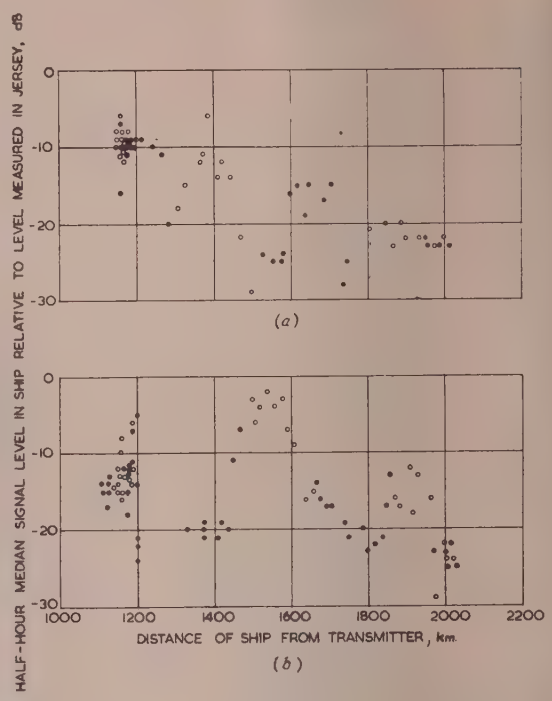


Fig. 12.—Variation of median signal level with distance (41 Mc/s).
(*a*) Outward voyage (7–10 October, 1953).
(*b*) Inward voyage (12–16 October, 1953).
○ Day
● Night

and receiving aerials and the nature of the foreground should be such as to permit effective low-angle transmission and reception.

Little significant difference in median values between day and night was noted in a series of observations at about 2 000 km, and with the long-range transmitting aerial in use the day-time overall median for all measurements at 1 900–2 050 km was only 1 dB above the corresponding night-time value. Similar results were obtained when the Jersey-directed aerial was in use, day and night median values then being about 2·5 and 3 dB respectively below those for the long-range aerial.

(7.2.2) Slough and Jersey Data.

Measurements were made simultaneously at both Jersey and Slough for eight periods of 2–4 weeks, spaced throughout 18 months. While the difference in length between the transmission paths from Shetland to the receiving stations was not great (1 185 and 935 km), these tests had the compensating advantage that the comparison of signal levels could be made over a longer period than for the ship trials. No attempt has been made to compare the median signal level for individual hours, but the median values of the hourly samples for each test period are compared in Figs. 6 and 7. It can be seen that the difference between levels at the two stations (ignoring the difference in median receiving-aerial gain of some 4–5 dB) was not great except during July and October, 1953, when it was 4 or 5 dB, but there is a tendency for the level at Jersey to be higher than that at Slough during periods of high signal and lower during periods of low signal.

(7.3) Beam-Spread Tests

(7.3.1) Ship Test Results.

The most important part of the beam-spread data was obtained as the ship steered a course forming approximately an arc of a circle centred on Shetland, by observation of the median signal

level when the Jersey-directed aerial was in use at the transmitter. Variations of the received signal level due to changing propagation conditions, as opposed to those due to angular displacement, were allowed for by comparing the median levels observed in the ship with those observed at the same time at Jersey; the results are given in Fig. 13. Unfortunately, the number of

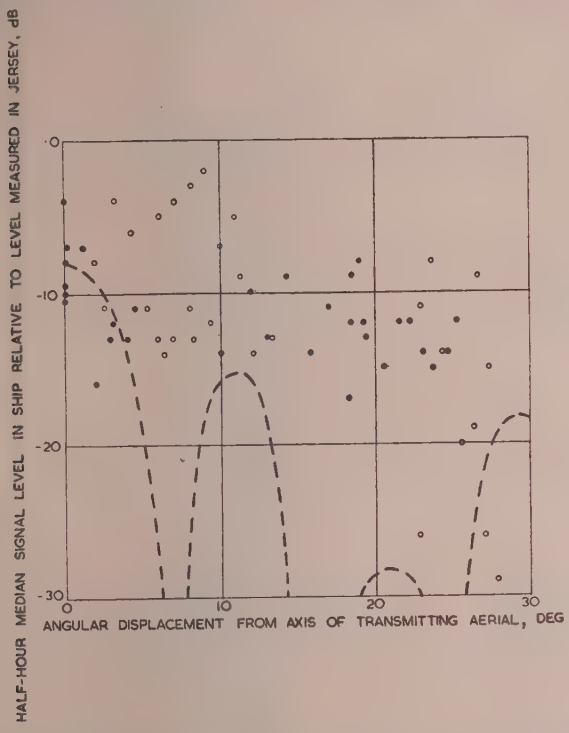


Fig. 13.—Variation of median signal level with angular displacement of receiving station from the direction of maximum radiation (41 Mc/s).

— — — Radiation diagram of transmitting aerial.
○ Day.
● Night.

measurements is barely adequate to enable a reliable comparison to be made between day and night conditions.

The received signal level decreases by only about 5 dB for off-course angles up to 20° and then decreases rapidly beyond 25°. When considering these results the radiation diagram of the transmitting aerial at the appropriate elevation, shown by the dotted curve in Fig. 13, should be borne in mind. In drawing this curve an allowance of 8 dB has been made for the relative effective aerial gains and feeder losses at Jersey and in the ship. It appears that even after allowing for the effect of side lobes of the transmitting aerial there is rather more scattering of radiation to points well off the great-circle path than would be expected on the theory of isotropic turbulence; it may be possible to account for some of this in terms of the theory of scattering from meteor trails put forward by Eshleman and Manning.[10]

During the range tests, observations of relative level from the Jersey-directed and long-range transmitting aerials showed very marked oscillatory fluctuations with distance, and there were times when the Jersey-directed aerial gave a higher signal than the aerial directed towards the ship. These aerials differed slightly in height above sea level and were on slightly different sites, so it is possible that the multi-lobed patterns due to sea reflections at the transmitting end were responsible for the complex results.

(7.3.2) **Slough and Jersey Data.**

Measurements of signal level were made at Slough and Jersey for transmissions from both the low-angle and Jersey-directed aerials during the whole time that the ship trials were in progress. This provided data on the beam spread over a longer period for a number of fixed off-beam angles. Jersey was 16° off-beam for the low-angle transmitting aerial, Slough was 6° off-beam for the Jersey-directed aerial and 22° off for the low-angle aerial. At both receiving stations a diurnal variation was observed in the relative median signal levels from the two aerials. This rose during the day-time and reached a maximum of 5–7 dB between 1600 and 2000 hours and fell to a low value of 1–2 dB during the night. The overall median difference in levels was 3 dB at Slough, and at Jersey was 3 dB at night, but 6 dB during the day. These differences are in general agreement with the trend of the measurements during the ship trials shown in Fig. 13, having regard to the difference in beam elevation of the two transmitting aerials and the difference between the beam widths for the receiving aerials at Slough and Jersey.

(8) MODULATION TESTS

The investigation of the use of scatter propagation for communication circuits described below was carried out over the Shetland–Jersey path, and included both telegraphy and telephony tests.

(8.1) Telegraphy

Part of the time during several series of tests was devoted to an examination of frequency-shift-telegraphy (f.s.t.) transmissions, both with and without wave-angle diversity reception. Standard 50-baud teleprinter signals were used, and the performance of the telegraph channel was assessed by counting the character error rate, precautions being taken to ensure that negligible distortion was introduced by the transmitting and receiving equipment. A block diagram of the telegraphy receiving equipment is shown in Fig. 14.

(8.1.1) **Bandwidth and Frequency Shift.**

Distortion in frequency-shift telegraphy transmitted by ionospheric scattering may arise from:

(*a*) Noise.
(*b*) Random frequency fluctuations of the order of 10 c/s (Section 4.1).
(*c*) Doppler frequency variations of up to 1 kc/s or more (Section 4.4).

In order to minimize the influence of noise a small bandwidth and frequency shift are desirable; however, the frequency must not be so small that random frequency fluctuations are a large fraction of the shift. These considerations suggest a frequency shift of about 100 c/s, but on the other hand, the relatively large frequency variations due to Doppler effects would suggest the need for a large frequency shift. Since it appears that those components of the signal having large Doppler frequency variations are often weak, it is not, in fact, necessary to use a shift in excess of 1 kc/s in order to reduce telegraph distortion due to this cause. Since the optimum value was not known it was decided to compare the performance obtainable with frequency shifts of 200 and 800 c/s.

A receiver half-power bandwidth of 900 c/s was used with the 800 c/s shift, and 300 c/s with the 200 c/s shift. Laboratory tests showed that a signal/noise ratio at the input to the f.s.t. demodulator of 10 dB with the 200 c/s shift and 9 dB with the 800 c/s shift was required for reasonably satisfactory operation (less than one error per 1 000 characters). With the 200 c/s shift system, calculation shows that a signal input to the v.h.f. receiver of −26 dB relative to 1 μV across the input is required to produce

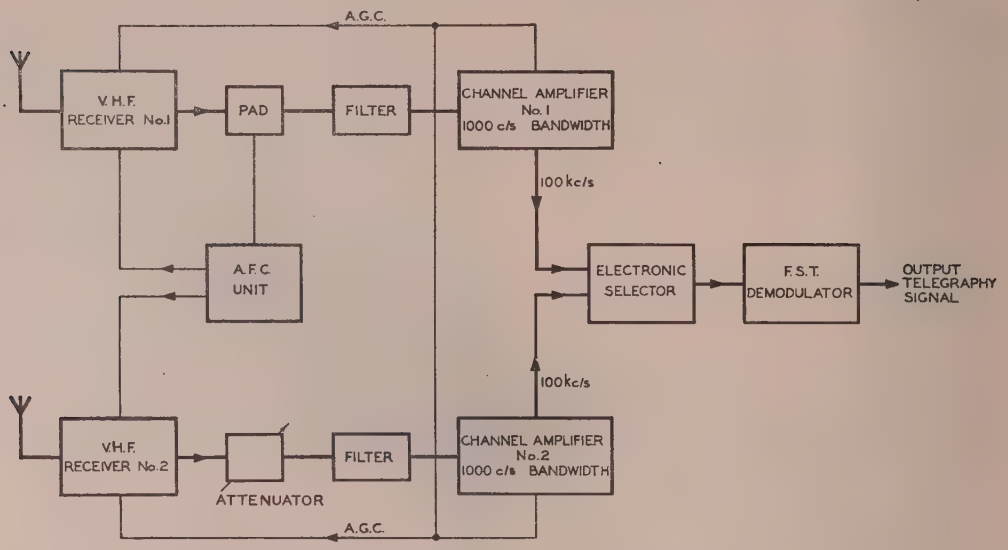


Fig. 14.—Block diagram of telegraphy receiving equipment.

a 10 dB signal/noise ratio in a 300 c/s band, assuming a receiver noise factor of 4 dB. With the 800 c/s shift system, with a 900 c/s bandwidth, the input signal level required to produce a 9 dB signal/noise ratio is −22 dB relative to 1 μV. These sensitivity figures were substantially confirmed in practice for steady signals.

The results of comparative tests between the two frequency shifts are shown in Fig. 15. It can be seen that the 200 c/s shift was better than the 800 c/s shift by a small margin. Other tests on a gradually dropping mean signal level confirmed that the 200 c/s system could be operated to a slightly lower level (i.e. the service maintained over a slightly longer period) than the 800 c/s system. The tests were carried out with diversity reception for both values of frequency shift.

Fig. 15 shows that the median input signal level for reasonably satisfactory operation should exceed about −10 dB relative to 1 μV; however, the telegraph errors observed on different occasions when the median signal level was at a given value below −10 dB relative to 1 μV varied over a wide range. It is seen that the median input signal level for satisfactory operation under conditions of scatter propagation is much higher (12–16 dB) than with steady signals.

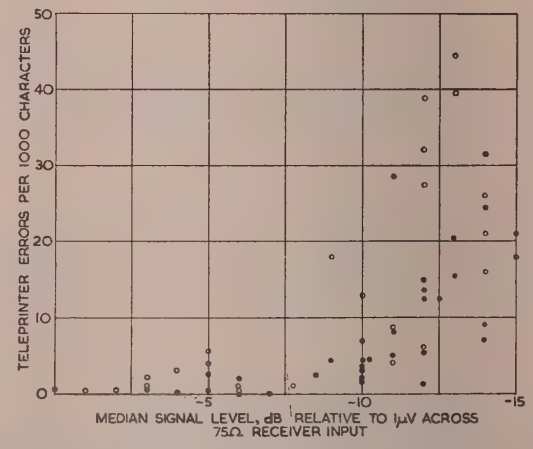


Fig. 15.—Variation of teleprinter errors per 1 000 characters with median signal level for shifts of 200 and 800 c/s (diversity operation on 41 Mc/s).

● 200 c/s shift.
○ 800 c/s shift.

(8.1.2) Diversity Improvement.

The results of comparative tests of single-aerial and two-aerial wave-angle diversity reception of f.s.t. signals are shown in Fig. 16 for 200 c/s shift. It can be seen that the improvement in error rate obtained by diversity reception is large, being generally greater than 10 : 1. There is considerable variation in the diversity improvement obtained at different times with the same median value of signal.

Observation of the signal envelope showed that, for median levels of −10 dB or above, the teleprinter errors were caused mainly by the short-duration fades described in Section 4.4 and were not directly attributable to the Doppler frequency variation associated with meteor reflections. When wave-angle diversity was used it was confirmed that these fades were usually displaced slightly in time as received on the two aerials, and the automatic selection by the diversity switch eliminated the effect of the majority.

(8.2) Telephony

During periods of reasonably high-level signals investigations were made into the transmission of telephony by phase modulation (p.m.), frequency modulation (f.m.) and single-sideband (s.s.b.) amplitude modulation. The full transmitter power of 4 kW was available for p.m. and f.m. tests, and for s.s.b. tests the peak sideband power was 3 kW. The transmitter was modulated in each case either directly from a normal telephone instrument or from a tape recording and reproducer. Peak clipping used in preparing the tape recording enabled the ratio of mean-to-peak sideband power in the s.s.b. tests, and the mean-to-peak deviation in the p.m. tests, to be made as large as possible.

The performance of the telephony channels was assessed by subjective listening tests. Tape recordings were made of some of the received speech signals, so that listening tests could be

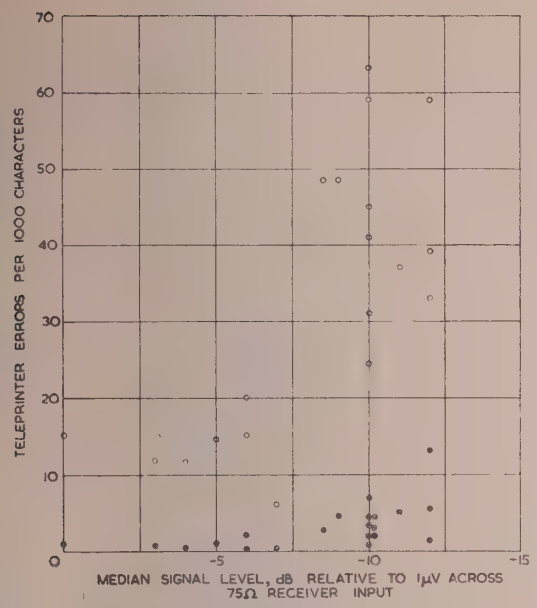


Fig. 16.—Comparison of teleprinter errors per 1 000 characters with and without diversity operation (41 Mc/s).

● With 2-aerial diversity.
○ Without diversity (average of two channels).

carried out later by different observers. Wave-angle diversity reception was not used. A block diagram of the telephony receiving equipment is shown in Fig. 17.

Preliminary tests showed that phase modulation was preferable to frequency modulation because of the less objectionable characteristics of the noise, and detailed investigation of an f.m. system was not pursued.

(8.2.1) **Phase Modulation.**

With phase modulation and an input of about 1 μV across the receiver input the received signal was almost completely intelligible, although the quality was marred by noise; a peak deviation of approximately 1·5 rad gave optimum results with a receiver half-power bandwidth of 5 kc/s. Although fairly frequent meteor whistles were heard, these did not seriously impair intelligibility. The main limitation to quality was the inadequate and fluctuating signal/noise ratio, the bursts of noise that arose when the received carrier amplitude fell below the noise level at the limiter input being particularly unpleasant.

(8.2.2) **Single-Sideband Amplitude Modulation.**

The s.s.b. system gave slightly better results than the p.m. system as judged by listening tests, the minimum acceptable median signal level input being about −2 dB relative to 1 μV. The noise was steadier (when an a.g.c. time-constant of a second or more was used) and the character of the noise was less unpleasant than with phase modulation; this may have influenced the subjective assessment of the performance.

Meteor whistles appeared to be less frequent and even less disturbing than with phase modulation, the intelligibility of the signal being limited by poor signal/noise ratio and not by whistles. The pilot carrier was radiated at a level 16 dB below the peak sideband power of 3 kW, and the absence of a high-level carrier probably did much to account for the reduction in meteor whistles. No difficulty was experienced in the operation of the receiver automatic frequency control from the pilot carrier, this satisfactory result being no doubt due to the selection of the pilot carrier in a filter of only 70 c/s bandwidth.

(9) POSSIBLE APPLICATIONS TO COMMERCIAL SERVICES

(9.1) Telegraphy

As pointed out in Section 8.1.2, a telegraph channel with an error rate generally well under 5 per 1 000 can be obtained with wave-angle diversity operation and a median value of the signal level not lower than −10 dB relative to 1 μV across the receiver input. This level at least must be maintained if such a circuit is to operate on a 24-hour basis throughout the year. It should be noted that with automatic error correction an error rate of 5 in 1 000 could be improved to at least 1 in 10 000.

The lowest 15 min median level observed in the 41 Mc/s tests

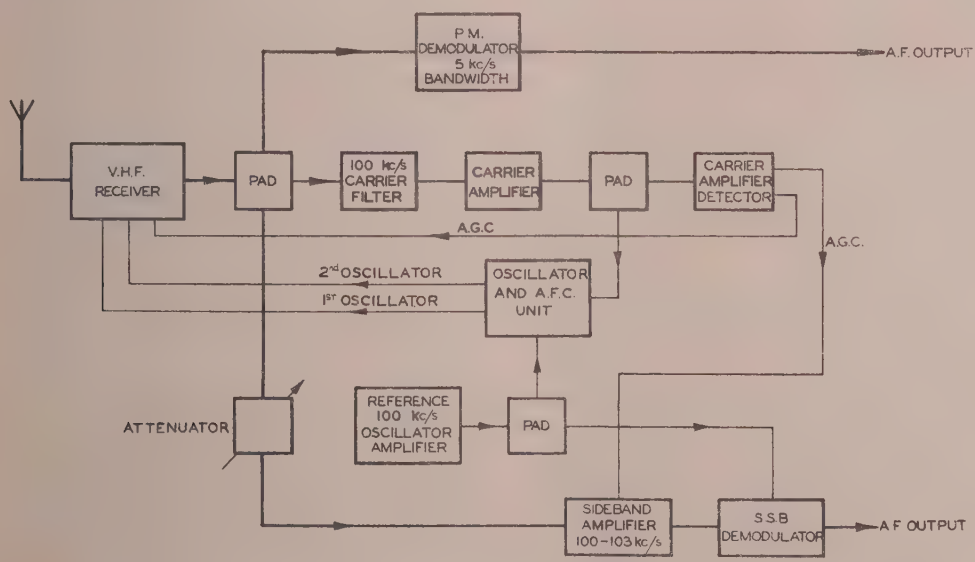


Fig. 17.—Block diagram of telephony receiving equipment.

was −28 dB relative to 1 μV and occurred in March, 1954. If this figure is taken as the lowest level likely to be encountered on a link such as that between Shetland and Jersey—and this is by no means certain, in view of the gaps between periods of observation—it would be reasonable to suppose that a satisfactory teleprinter service using the background-scatter signal could be established on a continuous basis by increasing the transmitted power and aerial gain by a total of about 18 dB relative to those used in the investigation. This could be achieved in practice by increasing the transmitted power to 60 kW (12 dB increase) and using a broadside array of four rhombics at each terminal (about 6 dB median improvement overall on scatter-type signals).

It is possible to provide a satisfactory telegraph circuit for a proportion of the time at much less expense. The proportion of the time for which a f.s.t. circuit is likely to be satisfactory at different times of year, and for various increases in effective power relative to the experimental system, may be estimated from Fig. 18. For example, an increase of only 10 dB would mean that the curves are moved 10 dB to the right. It can be seen that in midsummer such a circuit would give continuous service, and even in March it would achieve about 60% availability. A high proportion of the lost time would be during the night.

These performance estimates apply strictly to the experimental link investigated and to a frequency of 41 Mc/s. The estimates are probably applicable to other north–south links having lengths between about 600 and 1 200 miles and located at about the same distance from the auroral zone, but they would require modification for other frequencies and possibly for other path orientations.

Some improvement in performance may be expected by using frequencies somewhat lower than 41 Mc/s, but in view of the data obtained at 27 Mc/s (see Fig. 9) this improvement is not likely to be large; on the other hand, if frequencies appreciably higher than 41 Mc/s are used the system performance deteriorates rapidly.

(9.2) Telephony

If it is assumed that a received signal of −2 dB relative to 1 μV across the receiver input represents the lower limit for a tolerable telephony channel using single-sideband amplitude modulation, and that a 15-min median level of −28 dB is the lowest likely to be encountered, then an increase of about 26 dB in the received signal would be necessary for a continuous service. Such an increase is unlikely to be practicable, and would in any case be uneconomic.

As in the case of telegraphy, however, a telephony channel of limited availability is possible with appreciably less improvement relative to the system investigated than the 26 dB required for full availability. The percentage availability of a telephony channel at different times of the year, with various increases in effective power relative to the experimental system, may be estimated from Fig. 18. Thus an increase of 15 dB should give a circuit with an availability ranging from about 35 to almost 100% at various times of the year, much of the lost time being at night. It should be stressed again that care is required in applying these data to other links.

Although wave-angle diversity has not been applied to the reception of telephony signals with scatter propagation, it is probable that some advantage could be derived from its use. In particular, the short-duration bursts of noise in a p.m. system, or in a s.s.b. system having a short-time-constant automatic gain control, would be greatly minimized.

(9.3) Interference to and from Other Services

In the range of frequencies (25–60 Mc/s) which are technically

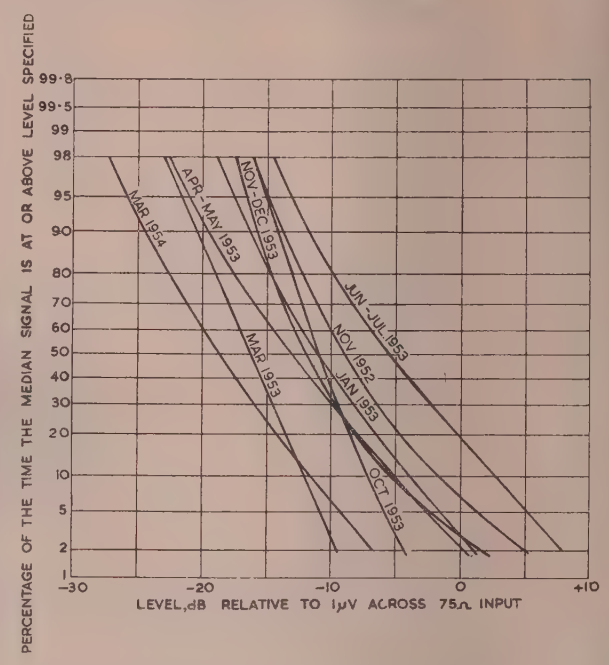


Fig. 18.—Distribution with time of hourly median levels of received signals at Jersey for each test period, November, 1952, to March, 1954 (41 Mc/s).

suitable for communication by ionospheric scattering, long-distance propagation by sporadic-E ionization may occur at times and give rise to interference between services sharing the same frequency band. At the low-frequency end of the range such signals may attain levels within 10–20 dB of those due to specular reflection, at distances up to some 2 000 km. However, the reflection coefficient falls rapidly with frequency, and for a given transmitter power the range of distances from which reception via sporadic-E ionization is possible falls steadily and becomes negligibly small for frequencies above about 60 Mc/s.[24]

Since sporadic-E ionization generally occurs in limited areas, it is unlikely that the wanted scatter signal and an interfering signal from a different direction would attain high levels at the same time. Under these conditions a weak scatter signal would be susceptible to interference from an occasionally much stronger signal from another transmitter propagated via sporadic-E ionization. Such interference could be reduced by the use of highly directional receiving aerials, or by the use of frequencies above about 45 Mc/s. It must also be remembered that, particularly in the years near to the sunspot maximum, long-distance propagation by the F_2 layer is possible in the frequency range 25–60 Mc/s.

Conversely, services using scatter propagation could themselves produce interference to other services using the same frequency band, if sporadic-E or intense F_2 ionization were prevalent. The foregoing observations on the choice of frequency to avoid interference also apply in this case; furthermore, the interference could be minimized by using highly-directional transmitting aerials for the scatter service.

(9.4) Fixed-Service Applications

From the point of view of possible applications to fixed-service links, a marked limitation of v.h.f. scatter-type propagation, as compared with normal h.f. propagation, is the upper limit of range of some 1 900 km. Greater distances could, of course,

be covered by a number of v.h.f. links in tandem, but these would inevitably be more expensive than a single-section h.f. link. On the other hand, v.h.f. scatter-type links offer greater reliability than h.f. links for telegraphy, provided that sufficient radiated power and aerial gain are used. It seems unlikely, however, that such links would be suitable for continuous operation of high-speed or multi-channel telegraphy, except perhaps for circuits carrying only a few time-division channels. While there is the possibility of exploiting a region of the spectrum not hitherto employed for long-distance services, it must be remembered that a large part of the frequency band 25–60 Mc/s which is technically suitable for v.h.f. scatter-type services is already in use in many countries for other services, including television, fixed and mobile services.

The main applications of v.h.f. ionospheric scatter-type links would thus appear to be

(*a*) To important telegraphy circuits where continuity of service is worth the extra cost.
(*b*) To suitable routes carrying heavy telegraphy traffic, to cover periods of ionospheric disturbance.
(*c*) To assist in the provision of frequency allocations for new telegraphy services.

The possible applications for telephony are less promising than for telegraphy in view of the very large radiated power required for continuous service; a limited field of application is possible when reliable communication in day-time only is sufficient.

(10) CONCLUSIONS

Tests carried out on the experimental links from Shetland to Jersey and Slough have provided valuable information on the general characteristics of v.h.f. ionospheric scatter propagation. It has been shown that a residual signal is always present, having amplitude and phase fluctuations indicative of scattering from a large number of highly variable scattering centres. It is probable that scattering due to both ionospheric turbulence and meteoric ionization contributes to this residual signal, the relative importance of these two mechanisms varying from time to time. Superimposed on the residual signal are numerous sudden enhancements or cancellations due to components from the ionized trails of the larger meteors, and a few longer-duration enhancements generally described as "sporadic-E propagation."

The random phase and amplitude fluctuations of the scatter signal and the low levels often experienced are such that this type of propagation is most suited to the transmission of signals of restricted bandwidth, e.g. telegraphy at 100 bauds or less and commercial telephony with bandwidths of not more than about 3 kc/s for distances up to about 1 200 miles.

The range of frequencies for which ionospheric scattering affords a useful means of communication is from about 25 to 60 Mc/s, but it should be borne in mind that a large part of this band is already in use in many countries for other services.

The effective gain obtained from directional aerials under conditions of scatter propagation approaches the plane-wave gain only for a small percentage of the time; the median value of the effective gain in a system with similar transmitting and receiving aerials is approximately equal to the square-root of the plane-wave power gain.

Telegraphy tests have shown that establishment of a 50-baud frequency-shift teleprinter service of substantially continuous availability is practicable, except perhaps during rare periods of auroral disturbance in the vicinity of the propagation path. A simple form of wave-angle diversity has proved to be of very great value in reducing the number of errors due to short-duration cancellation fades. However, it appears that continuous operation of high-speed single-channel telegraphy links, or multi-channel time-division telegraphy links, is unlikely to be satisfactory except perhaps for links carrying only a few channels.

A telephony system giving continuous service would necessitate an uneconomic or impracticable transmitter power, but a service having over 50% availability confined mainly to daylight hours is not impracticable. For telephony it appears that single-sideband amplitude modulation is to be preferred to phase or frequency modulation; the margin of advantage of single-sideband over phase modulation is small and both are appreciably superior to frequency modulation.

An outstanding advantage of scatter propagation is the relative immunity of such circuits to disturbance by ionospheric storms. The greater reliability of telegraphy circuits using scatter propagation, and the fact that frequency changes are unnecessary as compared with normal short-wave propagation, are further advantages.

(11) ACKNOWLEDGMENTS

This paper is published by permission of the Engineer-in-Chief of the General Post Office, the Director of Radio Research of the Department of Scientific and Industrial Research and the Lords Commissioners of the Admiralty.

The co-operation of the Admiralty in the ship tests is gratefully acknowledged.

The assistance given by the staff of the Post Office Laboratories at Dollis Hill and Castleton and of the Radio Research Station, Slough, in the erection of aerials, and the provision and operation of transmitting and receiving equipment was of great value and is acknowledged with pleasure; the authors are particularly indebted to Mr. R. Hamer for his substantial contribution to the work.

(12) REFERENCES

(1) ECKERSLEY, T. L.: "Short Wave Wireless Telegraphy," *Journal I.E.E.*, 1927, **65**, p. 600.
(2) ECKERSLEY, T. L.: "An Investigation of Short Waves," *ibid.*, 1929, **67**, p. 992.
(3) RATCLIFFE, J. A., and PAWSEY, J. L.: "A Study of the Intensity Variations of Downcoming Wireless Waves," *Proceedings of the Cambridge Philosophical Society*, 1933, **29**, p. 301.
(4) ECKERSLEY, T. L.: "Studies in Radio Transmission," *Journal I.E.E.*, 1932, **71**, p. 405.
(5) ECKERSLEY, T. L.: "Analysis of the Effect of Scattering in Radio Transmission," *ibid.*, 1940, **86**, p. 548.
(6) BAILEY, D. K., *et al.*: "A New Kind of Radio Propagation at Very High Frequencies observable over Long Distances," *Physical Review*, 1952, **86**, p. 141.
(7) BOOKER, H. G., and GORDON, W. E.: "A Theory of Radio Scattering in the Troposphere," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 401.
(8) MEGAW, E. C. S.: "Scattering of Electro-Magnetic Waves by Atmospheric Turbulence," *Nature*, 1950, **166**, p. 1100.
(9) MEGAW, E. C. S.: "Waves and Fluctuations," *Proceedings I.E.E.*, 1953, **100**, Part III, p. 1.
(10) ESHLEMAN, VON R., and MANNING, L. A.: "Radio Communication by Scattering from Meteoric Ionization," *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 530.
(11) MCKINLEY, D. W. R.: "Dependence of Integrated Duration of Meteor Echoes on Wavelength and Sensitivity," *Canadian Journal of Physics*, 1954, **32**, p. 450.
(12) NAISMITH, R.: "A Subsidiary Layer in the E-Region of the Ionosphere," *Journal of Atmospheric and Terrestrial Physics*, 1954, **5**, p. 73.

(13) ISTED, G. A.: "Atmospheric Electricity and Long-Distance Very High Frequency Scatter Transmissions," *Marconi Review*, 1954, **17**, p. 37.
(14) VILLARS, F., and WEISSKOPF, V. F.: "The Scattering of Electromagnetic Waves by Turbulent Atmospheric Fluctuations," *Physical Review*, 1954, **94**, p. 232.
(15) VILLARD, O. G., PETERSON, A. M., MANNING, L. A., and ESHLEMAN, VON R.: "Extended-Range Radio Transmission by Oblique Reflection from Meteoric Ionization," *Journal of Geophysical Research*, 1953, **38**, p. 83.
(16) VILLARD, O. G., and PETERSON, A. M.: "Meteor Scatter," *QST*, April, 1953, **37**, p. 11.
(17) EASTWOOD, E., and MERCER, K. A.: "A Study of Transient Radio Echoes from the Ionosphere," *Proceedings of the Physical Society*, 1948, **61**, p. 122.
(18) APPLETON, E. V., and NAISMITH, R.: "The Radio Detection of Meteor Trails and Allied Phenomena," *ibid.*, 1947, **59**, p. 461.
(19) LOVELL, B., and CLEGG, J. A.: "Radio Astronomy" (Chapman and Hall, London, 1952).
(20) BULLOUGH, K., and KAISER, T. R.: "Radio Reflections from Aurorae," *Journal of Atmospheric and Terrestrial Physics*, 1954, **5**, p. 189.
(21) MAXWELL, A.: "Turbulence in the Upper Ionosphere," *Philosophical Magazine*, 1954, **45**, p. 1247.
(22) SEDDON, J. C.: "Electron Densities in the Ionosphere," *Journal of Geophysical Research*, 1954, **59**, p. 463.
(23) SEDDON, J. C., PICKAR, A. D., and JACKSON, J. E.: "Continuous Electron Density Measurements up to 200 km," *ibid.*, 1954, **59**, p. 513.
(24) BRAY, W. J., HOPKINS, H. G., KITCHEN, F. A., and SAXTON, J. A.: "Review of Long-Distance Radio-Wave Propagation above 30 Mc/s," *Proceedings I.E.E.*, Paper No.1782 R, January, 1955 (**102** B, p. 87).

(13) APPENDICES

(13.1) Some Aspects of the Theory of Scattering

(13.1.1) Scattering due to Turbulence.

The important parameters in the theory of Booker and Gordon[7] are the scale of turbulence, l, which describes the inhomogeneity in the permittivity, ϵ, of the scattering medium, and $\overline{(\Delta\epsilon/\epsilon)^2}$. The troposphere is not significantly dispersive at radio frequencies, so that in this case $\overline{(\Delta\epsilon/\epsilon)^2}$ is independent of the wavelength, λ. The ionosphere, however, is dispersive, and it can be shown that $\overline{(\Delta\epsilon/\epsilon)^2} = \overline{(\Delta N/N)^2}\,(\lambda/\lambda_N)^4$. It then follows[6] that

$$\sigma(\theta, \chi) = \frac{\overline{(\Delta N/N)^2}(2\pi l/\lambda_N)^3 \sin^2\chi}{\lambda_N\left[1 + \left(\frac{4\pi l}{\lambda}\sin\theta/2\right)^2\right]^2} \quad . \quad . \quad . \quad (2)$$

In most of the experimental work described in the paper the angle χ in eqn. (2) does not differ much from $\pi/2$ and it is henceforth assumed that $\sin^2\chi = 1$.

The ratio of $(4\pi l \sin\theta/2)/\lambda$ to unity clearly has an important bearing on the variation of $\sigma(\theta)$ with λ. Thus, for given values of ΔN, N and λ_N, if $(4\pi l \sin\theta/2)/\lambda \gg 1$,

$$\sigma(\theta) \propto \left(\frac{\lambda}{\sin\theta/2}\right)^4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

but if $(4\pi l \sin\theta/2)/\lambda \ll 1$, then σ is independent of λ and θ. If the scattering is assumed to take place somewhere in the E-region, at heights, say, between about 80 and 120 km, then, for distances of transmission of the order of 1 000–2 000 km (probably the range of greatest interest in practice), θ will lie approximately in the range 20–30°. This means that the conditions above reduce respectively to $l \gg \lambda/2$, when $\sigma(\theta) \propto \lambda^4$, and $l \ll \lambda/2$, resulting in σ being independent of λ. In the region where l is of the same order as λ, the dependence of $\sigma(\theta)$ upon λ and θ is not expressible in terms of single exponents. Furthermore, in comparisons between the transmission characteristics at different frequencies it will be difficult to separate out changes in received power due to variations of $\sigma(\theta)$ with frequency alone, and variations of the radiation pattern of the scattering centres when l is not known precisely. Bailey and his colleagues[6] assumed a value of 100 metres for l, so that with their operating wavelength of 6 m the condition $l \gg \lambda/2$ would apply.

Let it be assumed that for a given transmission distance, D, the value of $\sigma(\theta)$ over the whole of the scattering region in the ionosphere of importance is constant and equal to that at the centre of the region. (For a given D the total range of variation of $\theta/2$ is unlikely to be more than about 3°.) Then, in transmission between identical transmitting and receiving aerials, each of effective aperture A and directed at the ionosphere (E-region) at the mid-point of the path, it can be shown[6] that the ratio of received to transmitted power is given by

$$\frac{P_r}{P_t} = \frac{4\sigma Ab}{D^2 \sin\theta/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where b is assumed to be smaller than the depth of overlap of the transmitting and receiving beams. For the same restriction on b it is shown in Section 13.2 that, if the transmitting and receiving aerials are not identical but have apertures A_t and A_r respectively,

$$\frac{P_r}{P_t} = \frac{4\sigma A_r b}{D^2 \sin\theta/2} \text{ for } A_t > A_r \quad . \quad . \quad . \quad . \quad (5)$$

and

$$\frac{P_r}{P_t} = \frac{4\sigma A_t b}{D^2 \sin\theta/2} \text{ for } A_r > A_t \quad . \quad . \quad . \quad . \quad (6)$$

On the other hand, if the scattering layer is relatively so thick that the effective volume for scattering is determined primarily by the region of overlap of the transmitting and receiving aerial beams,

$$\frac{P_r}{P_t} = \frac{4\sigma\lambda\sqrt{A_r}}{D(\sqrt{\pi})\sin\theta} \text{ for } A_t > A_r \quad . \quad . \quad . \quad (7)$$

and

$$\frac{P_r}{P_t} = \frac{4\sigma\lambda\sqrt{A_t}}{D(\sqrt{\pi})\sin\theta} \text{ for } A_r > A_t \quad . \quad . \quad . \quad (8)$$

It is thus clear that the received power does not depend uniquely on either A_t or A_r, and that the thickness of the scattering layer in relation to the overlap of the aerial beams—which may well vary from time to time—has an important influence.

The statistical theory of turbulent scattering advanced by Villars and Weisskopf[14] leads to conclusions somewhat different from those given above for the variation of $\sigma(\theta)$ with the wavelength. The theory considers a spectrum of eddy sizes extending down from l_0, which is a measure of the initial large-scale turbulence, to l_s, corresponding to the smallest eddy. When $\lambda \gg 4\pi l_s \sin\theta/2$, $P_r/P_t \propto \lambda^{13/3}$; and when $\lambda \ll 4\pi l_s \sin\theta/2$, $P_r/P_t \propto \lambda^{11}$. There should therefore be a rapid variation of the exponent of λ in the expressions for $\sigma(\theta)$ when λ is of the same order as $4\pi l_s \sin\theta/2$.

Before going on to describe briefly the characteristics to be expected of signals forward scattered by meteor trails, it should be observed that Eckersley[4] developed a theory of ionospheric scattering processes on the hypothesis of spherically symmetrical clouds, the radii of which were supposed large compared with the wavelength. On this basis, and ignoring any effects of

absorption (which would become more important with increasing wavelength), he showed that the scattered power in a given direction should be proportional to $[\lambda/(\sin\theta/2)]^8$, provided that the frequency concerned is well in excess of the plasma frequency corresponding to the ionization density of the cloud.

(13.1.2) Scattering from Meteor Trails.

It has been suggested by Eshleman and Manning[10] and McKinley,[11] on the basis of approximate calculations, that the observations of Bailey and his colleagues[6] over the Cedar Rapids–Sterling (U.S.A.) transmission path can be accounted for just as well on the assumption of scattering from ionized meteor trails as by the theory of turbulence outlined above.

There has existed for some time a considerable body of knowledge* on the properties of back-scattered echoes from meteor trails, and it is certain that meteoric ionization is an especially important source of non-uniformity in the E-region. Work at Stanford University (U.S.A.)[15,16] indicates that reflections from meteor trails can support, intermittently, high-frequency radio transmission well beyond the line of sight without the aid of normal reflections from the E- or F-layers. It has been shown that the duration of a forward-scattered signal from a given trail can be many times that of a back-scattered echo; this effect should therefore greatly increase the percentage of time for which forward transmission can be supported as compared with the duration of signals scattered back from meteoric ionization. The Stanford workers[10] assumed that each meteor trail is a long, thin, ionized cylinder lasting for a time of up to about one second, and for such trails the scattering is markedly dependent upon the orientation of the trail in relation to the transmission path. McKinley includes consideration also of longer-duration trails for which there is much less, and in some cases no, aspect sensitivity.

There is no need here to elaborate the details of the theoretical calculations; there are in any case minor differences in the results obtained depending upon the assumptions made in the treatment. It is clear, however, that the scattered power to be expected varies as λ^n, where n is of the order of 6—a value also obtained, it may be noted, by Eastwood and Mercer[17] from a study of radar echoes in the range 22–44 Mc/s. Furthermore, the scattered power also decreases as θ increases. It is not certain precisely how the received power in given circumstances should vary with the characteristics of the transmitting and receiving aerials. For the conditions assumed by Eshleman and Manning, in which randomly oriented trails are postulated, it appears that the received power would not depend upon the aerial gain to any significant extent. On the other hand, as McKinley points out, meteor trails are not orientated randomly, certainly not in their spatial orientations, and possibly not in time. The distribution of the known meteor radiants is very complex; and a further uncertainty is that meteors contributing to forward scattering may be several magnitudes smaller than those which have so far been systematically studied (by back-scatter and visual techniques), and their distributions are unknown. McKinley's work suggests that the full gains of the transmitting and receiving aerials might be realized. There is no complete unanimity of view about the height in the ionosphere at which most meteor trails occur, and it appears that all heights in at least the range 80–120 km are possible.

In practice, when making use of scattering from meteor trails it would probably be best to direct the transmitting and receiving aerial beams to one side or the other of the great-circle path by a few degrees, as suggested by Eshleman and Manning. In this way a greater number of suitably orientated trails will be available from meteors with the higher radiants, although, as pointed out by McKinley, scattering from the longer-duration trails is not affected to the same extent by this argument. This particular feature of meteoric reflections might be used to assess the relative importance of scattering due to meteor trails and that due to turbulence.

(13.1.3) General Dependence of Scattering on Frequency and Aerial Characteristics.

From the foregoing discussion it will be evident that, with so much uncertainty concerning the important parameters in the two general theories, it is hardly profitable at present to attempt too quantitative a comparison between them and experimental results. In so far as it is possible to express simply the ratio of P_r/P_t over a given transmission path in terms of frequency, aerial characteristics and scattering angle, it seems that all that can be expected is approximate conformity with a relation such as

$$\frac{P_r}{P_t} \propto \frac{A_t^n A_r^p f_N^r}{f^m (\sin\theta/2)^q} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Since, from a practical aspect it seems best to make a comparison between the performance at two different frequencies over any given link for systems having the same transmitting aerial gains, G_t, at the two frequencies, and also the same receiving aerial gains, G_r (under which conditions the same scattering volume would be concerned in the two cases), it is convenient to rewrite relation (9) as

$$\frac{P_r}{P_t} \propto \frac{G_t^n G_r^p f_N^r}{f^s (\sin\theta/2)^q} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where n and p both lie in the range 0–1 (but are not necessarily equal). If turbulent scattering is the dominant mechanism it follows from the discussion presented in Section 13.2 that, to a first approximation, if $G_r > G_t$ then $p = 0$ and $\frac{1}{2} \leqslant n \leqslant 1$, while if $G_t > G_r$, $n = 0$ and $\frac{1}{2} \leqslant p \leqslant 1$; under the same conditions $r = 4$. The factor involving the plasma frequency, f_N, occurs specifically only in the theory of turbulent scattering; it should be noted, however, that it has been suggested by Appleton and Naismith[18] that the susceptibility of the ionosphere to ionization by meteors probably increases as the existing density of ionization in the E-region increases. For meteoric ionization, therefore, some dependence of P_r/P_t on f_N may be expected (i.e. $r > 0$), but its precise nature is unknown.

The values of the exponents s and q in relation (10) are even less certain, i.e. purely on theoretical considerations. For turbulent scattering, and following the modified Booker–Gordon theory, s may lie in the range 2–6, corresponding to the range 0–4 for σ (see above), while q might take values 1–5, both s and q depending upon the ratio of the scale of turbulence to the wavelength. If the Villars–Weisskopf theory more nearly represents the actual scattering process, values of s and q considerably greater than 6 and 5 can conceivably occur: s and q also depend upon the relative importance of the contributions to the received signal due to turbulence and meteoric effects in any given instance, although it seems that if scattering from meteor trails is the dominant mechanism s may be expected to be of the order of 6. The factor $(\sin\theta/2)^q$ is not completely adequate to express the angular dependence of scattering from such trails, for an additional factor involving the angles between the planes of transmission and the directions of the trail axes should be included. In the experiments described in the paper attention is chiefly focused on the exponents n, p and s, comparison of transmission characteristics being made for fixed ranges and thus for substantially fixed values of θ. A word of caution must,

* This is well summarized in a recent survey article "Radio Echo Studies of Meteor Ionization," by T. R. Kaiser, *Advances in Physics*, 1953, 2, p. 495. See also "Meteor Astronomy," by B. Lovell (University Press, Oxford, 1954).

however, be injected in this connection, for if the scale of turbulence is of the same order as the wavelength, the dependence of P_r/P_t on f and θ will in any case be of a rather more complex nature than that indicated by relation (10).

Finally, in none of the theoretical discussion so far has the possible effect of absorption in the ionosphere been mentioned, but it should be pointed out that, at the lower end of the v.h.f. band, such absorption may be great enough to influence the effective value of s. This matter is discussed in more detail in Section 5.3 in relation to the experimental results.

(13.2) Dependence of Received Power on Aerial Characteristics and Thickness of Scattering Medium

(13.2.1) Thin Scattering Layer.

First let it be assumed that the extent in height of the scattering layer, b, is small compared with the depth of overlap of the transmitting and receiving aerial beams. No first-order error will be made if it is assumed that the layer is plane.

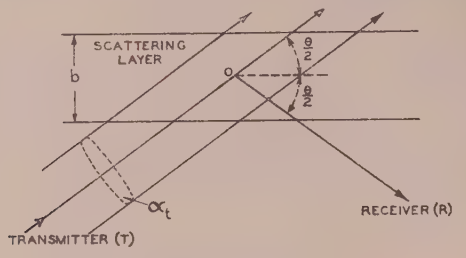


Fig. 19.—Scattering by a thin layer.
OT = OR = $D/2$.

Consider the situation illustrated in Fig. 19, and suppose that the transmitting aerial has an effective aperture A_t, and gain G_t, with respect to an isotropic radiator.

Thus
$$G_t = \frac{4\pi A_t}{\lambda^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Let it also be assumed that the radiated beam is conical in overall shape, and of semi-angle β, so that $\beta = \lambda/\sqrt{(\pi A_t)}$ and the solid angle of the beam is $\pi\beta^2$ or λ^2/A_t. If α_t is the cross-sectional area of the beam at the mid-point, O, of the scattering volume (where $\alpha_t = \lambda^2 D^2/4A_t$), the incident power density at O is P_t/α_t. Since β will generally be of the order of a few degrees and $b \ll D/2$, it can be assumed that the power density is constant over the effective scattering volume, and that this volume, V, is approximately equivalent to that of a cylinder of cross-sectional area $\alpha_t/(\sin\theta/2)$ and height b; thus $V = \alpha_t b/(\sin\theta/2)$.

The contribution, ΔP_r, to the received power at R due to an element of volume ΔV at O, ignoring any effects of absorption in the ionosphere, will be

$$\Delta P_r = \frac{P_t}{\alpha_t}\frac{4A_r}{D^2}\sigma(\theta)\Delta V \quad . \quad . \quad . \quad . \quad (12)$$

since, if the receiving aerial has an effective aperture A_r, the solid angle subtended by this aperture at O is $4A_r/D^2$. The total power received is therefore

$$P_r = \frac{P_t}{\alpha_t}\frac{4A_r}{D^2}\oint_V \sigma(\theta)dV \quad . \quad . \quad . \quad . \quad (13)$$

where $\oint_V$ indicates integration over the whole of the volume V, and it is assumed that A_r/D^2 may be regarded as constant over the scattering volume. There is also the further important proviso that the whole of this volume V falls within the beam of the receiving aerial, which implies that $A_t > A_r$, or $G_t > G_r$.

Since in most practical cases θ is of the order of 25°, and since the variation of θ over the volume is not likely to differ by more than one or two degrees, it may further be assumed that $\sigma(\theta)$ is constant over V (and equal to the value at O).

Therefore
$$P_r = \frac{P_t}{\alpha_t}\frac{4A_r}{D^2}\sigma V$$

or
$$\frac{P_r}{P_t} = \frac{4\sigma A_r b}{D^2 \sin\theta/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$$= \frac{\sigma\lambda^2 G_r b}{\pi D^2 \sin\theta/2} \text{ for } A_t > A_r \quad . \quad . \quad . \quad (15)$$

A similar argument shows that, when the receiving aerial is of greater directivity than the transmitting aerial ($A_r > A_t$), and when the effective scattering volume is therefore limited by the intercept of the receiving aerial beam with the layer,

$$\frac{P_r}{P_t} = \frac{4\sigma A_t b}{D^2 \sin\theta/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

$$= \frac{\sigma\lambda^2 G_t b}{\pi D^2 \sin\theta/2} \text{ for } A_r > A_t \quad . \quad . \quad . \quad (17)$$

(13.2.2) Thick Scattering Layer.

Suppose now that the thickness of the scattering layer is greater than the depth of overlap of transmitting and receiving aerial beams, as illustrated in Fig. 20. To a first approximation

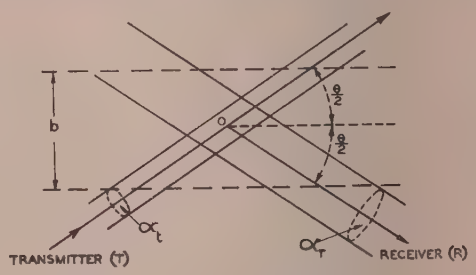


Fig. 20.—Scattering by a thick layer.

the beams, if roughly of pencil form, may be represented in the region of overlap by cylinders of cross-sectional area α_t and α_r, the values respectively at O, the mid-point of the scattering region.

If $A_t > A_r$ (i.e. $\alpha_t < \alpha_r$), the volume common to the two beams, which will therefore be the effective scattering volume, is given approximately by $\frac{\alpha_t}{\sin\theta} 2\sqrt{\frac{\alpha_r}{\pi}}$. Assuming once more that $\sigma(\theta)$ can be regarded as constant and equal to the value at O and again ignoring absorption, an argument similar to that of the preceding section shows that

$$P_r = \frac{P_t}{\alpha_t}\frac{4A_r}{D^2}\sigma\frac{\alpha_t}{\sin\theta}2\sqrt{\frac{\alpha_r}{\pi}} \quad . \quad . \quad . \quad . \quad (18)$$

or, remembering that $\alpha_r = \lambda^2 D^2/4A_r$,

$$\frac{P_r}{P_t} = \frac{4\sigma\lambda\sqrt{A_r}}{D\sqrt{(\pi)}\sin\theta} \text{ for } A_t > A_r \quad . \quad . \quad . \quad (19)$$

$$= \frac{2\sigma\lambda^2\sqrt{G_r}}{\pi D \sin\theta} \text{ for } A_t > A_r \quad . \quad . \quad . \quad . \quad (20)$$

Similarly, if $A_r > A_t$ (i.e. $\alpha_r < \alpha_t$)

$$\frac{P_r}{P_t} = \frac{4\sigma\lambda\sqrt{A_t}}{D\sqrt{\pi}\sin\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$= \frac{2\sigma\lambda^2\sqrt{G_t}}{\pi D\sin\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

While it is obvious that the assumptions concerning the shapes of the aerial beams and the constancy of $\sigma(\theta)$ in the foregoing analysis must impose some limitations on the quantitative applicability of the relations (14)–(22), there seems little doubt that a definite difference between the performance of the aerials may be expected depending upon whether the layer is thin or thick in relation to the depth of overlap of the beams.

DISCUSSION BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 31ST OCTOBER, 1955

Dr. E. C. S. Megaw: The paper is a valuable contribution to our factual knowledge and tackles one or two interesting problems which have not been closely touched before. The Introduction is particularly praiseworthy: it starts at the beginning of the subject of scattering of relatively short radio waves in the work of T. L. Eckersley, gives that work its proper recognition and then goes on to state clearly and fairly how the current possibilities of scattering of one kind or another from the E-region for practical, moderately long-distance communications came to be realized.

In the body of the paper I would point particularly to the experiment—perhaps the first convincing and complete one of its kind—on the lateral spread of the signal. The results give interesting support to the conclusion that the scattering arises in part from meteor ionization and in part from turbulent fluctuations, so helping to resolve what has seemed at times an unduly parochial argument.

The discussion of frequency dependence I would count to the credit of the paper, although I have some criticism in detail. The demonstration that absorption in the lower part of the E-region is likely to be of practical significance at about 30 Mc/s is interesting. I agree that it is difficult to think of anything else that would make the curve illustrating the frequency dependence flatten to the degree it does in this region. The explanation of the reduced effective aerial gain, to something like the square root of the plane-wave gain, in terms of the difference between a relatively thick layer and the relatively thin layer considered by Bailey *et al.* is also noteworthy. It is perhaps doubtful, however, whether the effective layer thickness is likely to be great enough to account for the whole of the reduction in gain.

My only substantial criticism of the paper relates to the more detailed interpretation of the frequency dependence, and particularly to some points in the discussion in Section 13.1. In this respect the authors' start from the empirical equation (1) was not very good, as they recognize later in the paper. It is perhaps worth emphasizing that there was never any sound reason to expect the frequency dependence to follow a simple power law. In Booker and Gordon's formulation of the problem [eqn. (2)] it is true that, in the relevant circumstances ($l \gg \lambda/2$), one finds a simple power-law dependence. This is merely a statement of the well-known fact that the refractive index of the ionized medium varies as the inverse square of the ratio of the frequency to the (much lower) critical frequency, while the scattering process itself is independent of frequency. But this is a consequence of quite specific assumptions made about the refractive-index fluctuations and, by implication, about the turbulent medium in which they are supposed to arise. In particular, it is a consequence of the assumption that the fluctuations can be described by an exponential correlation coefficient. As I pointed out on an earlier occasion,* this amounts to saying that the fluctuation spectrum of "eddy sizes" follows the same law down to eddies of infinitesimal size. This cannot occur in Nature; the energy dissipation in the turbulence would be infinite if it did. Whether, or how much, this matters depends on the conditions of the problem (wavelength, scattering angle, etc.), but especially on the smallest spatial size of the fluctuations actually present, and so, via the pressure, on the height in the atmosphere. In tropospheric scattering it happens that it does not matter (or at any rate not much), because the real spectrum does extend to eddy sizes so small that if it extended to still smaller ones it would make little difference. In the E-region it is quite otherwise: the smallest eddy size, no more than a centimetre in the lower troposphere, has grown to roughly 10 m. Thus, as Villars and Weisskopf have pointed out (Reference 14 of the paper), the increasingly rapid fall-off of scattered power observed in the 50–100 Mc/s region is just what one would expect. In the original Booker–Gordon formulation, however, such a possibility is not even contemplated; the success of this formulation in predicting the right order of magnitude for the ionospheric scattered field can fairly be described as an accident, although historically a happy accident.

In the paper there are a number of statements about the relationship between the wavelength and the scale of turbulence, and it is not always clear whether what is meant by "scale of turbulence" is the "eddy size," or the "integral scale" l, which is a length representing in order of magnitude the biggest eddy size, or something different. This confusion seems to have been responsible for some difficulty in appreciating the simple significance of the important conclusion of Villars and Weisskopf just mentioned.

Dr. R. L. Smith-Rose: Apart from one or two comparatively minor contributions elsewhere, the paper is the first published in this country to describe a comprehensive series of tests indicating the possibilities of communication by the scattering of radio waves from the ionosphere.

Those who have been studying the physics of the ionosphere for the past 25 years are interested in seeing what the engineer can do when he uses sufficient power to demonstrate the properties of scattering from the ionosphere. While he is not quite able to understand the causes of this phenomenon, the paper provides the scientist with additional material which should encourage him to continue his investigations into the characteristics of the ionosphere in more detail.

The attempt to confirm the relation between frequency and efficiency of transmission does not give a simple answer, and it is suggested, therefore, that there may be two or more causes which result in the overall propagation by scattering. One of these seems to be connected with meteors in the atmosphere, and I was intrigued by the fact that whistles can be heard in background noise which is ascribed to these meteors. It is not clear why one should hear an audible whistle as the result of interference between two coherent effects of radiation.

I would add a word of caution on the engineering side: the subject is topical, and much experimental work has been done; most of it has not been disclosed until recently, and I think that we are in danger of the whole possibilities of the system being oversold. Although it has been shown that communication is possible over a limited range of distance—1 000 to 1 200 miles—when the facts are investigated it appears to be a very extravagant and

* *Proceedings I.E.E.*, 1953, **100**, Part III, p. 1.

inefficient system, and it seems inconsistent with the general principles of engineering development to go backwards in the matter of efficiency.

From a study of factors which must be borne in mind when considering the development of this system and its exploitation on any considerable scale, it is clear that the transmitter power used is much higher than is normal in a radiocommunication system. While this power must necessarily be concentrated in a certain direction, it is not easy to design aerial systems to make sure that the radiation is going where it should, and radiation in other directions and even behind the reflector may be comparable with the power radiated by normal transmitters used for communication over much larger distances. The possible resulting interference must be very carefully considered.

Secondly, the efficiency of the aerial system is also poor, the gain being of the order of the second root of the theoretical value. Finally, it appears that the receivers are inefficient and their sensitivity is 10–12 dB below that of receivers used for the normal reception of coherent-signal transmission.

Frequency allocation is a subject with which the radio engineer is much concerned at present, because there are no spare frequencies in this portion of the spectrum. As is shown, the system would appear to work over the range 25–60 Mc/s, and there seems to be some advantage in keeping the frequency as low as possible. On the other hand, all the development and experimental work on this subject during the past 3–4 years has been carried out during a period of minimum sunspot activity; thus, during the next few years of increasing solar activity the frequencies used by normal high-frequency communication services will tend to rise above 30 Mc/s. In fact, during the last period of maximum sunspot activity, frequencies of 50 Mc/s and higher were used for transmission over long distances. This is something to which thought must be given by those who will plan the future of communication services using this scattering type of transmission. As the existing service frequencies increase they will suffer interference from any station in the 25–40 Mc/s bands, and I think that the receiver which uses scattering transmission may also be subject to some interference from normal high-frequency transmissions.

Dr. E. V. D. Glazier: The economics of this system of communication are marginal, and the transmitter power and aerial gains are such that every extra decibel of real gain costs many thousands of pounds. The designer must therefore have propagation information of the highest order of accuracy, in spite of the statistical nature of the signal.

I have been monitoring a link somewhat longer than the Shetlands–Jersey one, and when the transmitter is corrected to the power used by the authors I have a median signal over the best months of the year which compares well with Fig. 18. The transmitting aerial had 14 dB plane-wave free-space gain, and I believe it was not properly sited in relation to the correct angle of beam elevation. The receiver aerial had only 5 dB gain and was correctly sited. From the authors' rule for aerial gain degradation it would appear that this alternative information is almost 10 dB more optimistic than the authors'. However, in order to make a more exact comparison, I should like to know whether the reflecting zone ahead of the authors' aerials was sufficient to ensure the full 6 dB site gain for the quoted angle of beam elevation, and do the rhombic aerial gains given in Table 1 include this 6 dB gain due to a combination of the direct and reflected waves? Do the authors anticipate any latitude effect on the propagation?

Seeking points where a few more decibels of real gain might be found, one examines the bandwidth; it would appear that one teleprinter requires a bandwidth of only about 70 c/s (not the 200 c/s used by the authors), and this could be reduced even further by synchronous operation with start and stop signals omitted. A narrow-band receiver of this type would need special arrangements—such as the use of two separated filters—to overcome meteor whistles.

In their frequency-dependence law the authors do not mention the change of cosmic noise with frequency, the usually accepted law being $1/f^{2.41}$. Thus, from 30 to 60 Mc/s the noise voltage drops by almost the same amount as the signal median value, and on this argument Fig. 9 shows that frequencies from 30 to 60 Mc/s are equally good from the propagation aspect. If this is correct, the higher range of frequencies is preferable owing to reduced sporadic-E and F-layer interference. However, experience seems to indicate a preference for frequencies below 40 Mc/s.

The type of signal-level distribution shown in Fig. 4 indicates that for error rates better than about 10% every 3 dB real gain in the link halves the error rate. There comes a point when 3 dB more transmission power or real aerial gain costs £20 000 or more. For this sum an automatic error-correction telegraph equipment can probably be provided and supplied with the necessary skilled maintenance. When the signal/noise ratio is good the multi-path effects are relatively small, and quite high signalling speeds can be used: several hundred bauds are possible, so that eight time-division channels might be operated. It should not be impossible to devise a system which operated at that speed when the signal was good, and which by means of the automatic error detecting and correcting principle caused the system to stop working during fades without printing errors.

In view of the probability of interference in the area in front of the transmitting aerials, there is a case for siting these stations on the coast.

The system can make a real contribution to the relief of congestion in the h.f. band, and there may prove to be a case for having a small amount of bandwidth around 40 Mc/s cleared for it; two bands of about 100 kc/s separated by about 2 Mc/s would probably suffice for an appreciable number of links.

Mr. R. Hamer: It is stated in Section 4 that the received signal was always present, and I think that may be misleading. If by "present" it is meant that the signal could be detected in the presence of the receiver noise, then the signal was "absent" in the January, 1953, tests, for example, for about 2% of the time. This is in addition to the rapid cancellations in the received signal referred to in Section 4.1.

I would say that the diversity signal records of Fig. 3 show, not a lack of correlation, but a form of delayed correlation which accounts for the much improved reception of telegraphy signals using diversity.

It is not clear what Fig. 11 is intended to show. Since the gain is plotted against the median signal as received on the Yagi array, this median level will naturally tend to be high when the effective gain (relative to the dipole) is high. Is it not more appropriate to plot the effective gain against the median level of the signal received on the dipole? Measurements made at Jersey have been plotted in this way, and they showed that the gain of a rhombic aerial was often higher when the received signal level was lower. The results of all such measurements showed no significant correlation between the effective gain of a rhombic and the signal level received on a dipole. These observations do not apply to burst signals or sporadic-E reception, when higher effective gains were measured, as expected.

In view of the variation in the height, and possibly the depth, of the dominant scattering region, do the authors agree that it is preferable to increase the gains of the aerials, beyond a certain value, by increasing the directivity in the horizontal rather than in the vertical plane, unless the aerials are steerable in the vertical plane?

Mr. E. Fitch: The phenomenon of gain degradation is of great

importance in the exploitation of scatter from the ionosphere. Experiment has shown that an aerial of plane-wave power gain G acts rather as if it had a power gain of $\sqrt{G}$ when transmitting or receiving scatter-type signals. The authors have suggested a model in Section 13.2 which accounts for this degradation at one terminal but not at the other. I should like to suggest a modification of the model which at the same time describes the experiment more closely and leads to gain degradation at both aerials. The authors' model requires that the aerials have pencil beams, but the angle between the centre of the beam and the horizontal is usually less than the beam width; in fact, it seems reasonable to choose the height of the aerials so that the direct ray and the ray reflected from the ground in front reinforce each other in the direction of the ionosphere at the middle of the path. The effect is to reduce the vertical beam-width of the aerial below that calculated on the basis of the aerial gain in about the ratio $4\beta/\theta$. Calculation then gives, for a thin scattering layer,

$$\frac{P_r}{P_t} = \frac{4\sigma\lambda^2\sqrt{(G_t)G_r b}}{\pi D^2} \text{ for } G_t > G_r$$

and for a thick layer

$$\frac{P_r}{P_t} = \frac{\sigma\lambda^2\sqrt{(G_t G_r)}}{\pi D}$$

These formulae include the gain at the transmitter and receiver due to coherent addition of the direct and reflected rays.

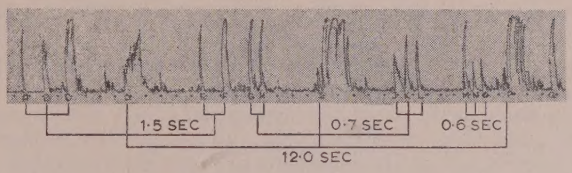


Fig. A

Fig. A is a record of the Shetlands c.w. transmission received at Great Baddow, Essex—a path length exceeding 1 000 km. In this record the vertical excursions correspond to increases in amplitude of the signal of the order of 20 dB from the base line to the saturation point of the equipment; the time scale, in seconds, is indicated just below the base line. Each burst of signal has been labelled A, B, C, etc. It will be seen that the signal bursts A, B and C are equally spaced; this spacing is reproduced again between E and F. On the other hand, the spacing between G and H is the same as the spacing between J, K and L. Other signal bursts in this record which exhibit regularity of spacing can be found in M, N and O, and in D, I and P, the spacing in the latter group being as much as 12 sec. Q is the only major signal burst, of the 17 shown, which cannot be time-related to the other signal bursts shown. It seems rather fantastic to believe that these bursts can be so arranged in time if they are produced by meteors.

Fig. B shows three records taken within minutes of each other

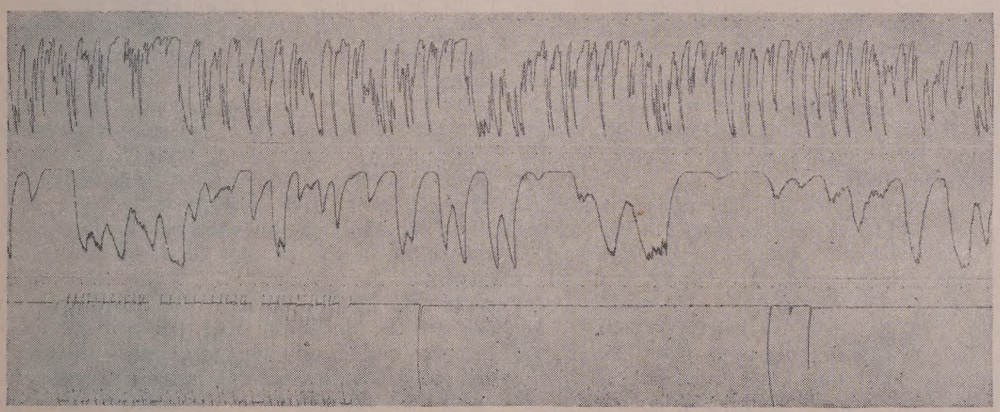


Fig. B

If one is obliged by the nature of the terrain to have aerials at different heights, it may be advantageous to employ aerials of different gain so that their beams may overlap more effectively.

There is a danger in drawing curves such as Fig. 11, for the ratio (or difference) of two uncorrelated variates may well be correlated with both. If the Figure is replotted to show the gain as a function of the signal received on a dipole, little correlation is to be seen. It is more informative to plot the signal received on the array as a function of the signal received on a dipole; then one finds that the signals are strongly correlated, with the exception that those from the dipole never fall below about 25 dB below 1 μV in 75 ohms. Reference to Fig. 8 suggests that this may be due to the high noise level at the receiving station. It may be untrue to say that aerial gains are more likely to be achieved when the signal level is high.

Mr. G. A. Isted: The authors stress the important contribution made by meteor ionization to scattered v.h.f. signals, but I wonder whether we are being too easily persuaded that the effects noticed are really due to meteors.

of the Shetlands transmission. The time scale, in seconds, is again shown as dots on each strip. The top strip is typical of the normal scatter type of signal. The second strip shows that the character of the signal has undergone a definite change; this change was not brought about instrumentally. The bottom strip shows that the signal has undergone yet another change and now has the character of strong sporadic-E ionization. The station call sign was being transmitted at the beginning of this strip. This gradual transition from the scattered signal to the sporadic-E condition has been recorded on numerous occasions, and it suggests that the mechanism responsible for the scatter signal may, in some way, be associated with the mechanism which produces sporadic-E ionization. So far as I know, however, there is no evidence to suggest that meteors have any influence on the formation of sporadic-E ionization.

Mr. J. P. Titheradge: Have the authors considered the possible application of combined space and frequency diversity transmission to long-range communication by ionospheric scattering? With amplitude modulation the transmitters, modulated simul-

taneously in phase, would have their carrier frequencies spaced by, say, 10–15 kc/s. The resulting overall improvement in the combined received signal might enable considerably lower individual transmitter powers to be employed.

As stated in the paper, the investigations show that improvements can be obtained from both wave-angle and frequency-diversity reception. It would seem then that these advantageous effects could also be produced at the transmitting end to afford a possible economy in transmitter cost, i.e. two low-power transmitters might cost far less than one of, say, 10 dB higher rating. Similar economy might also be obtained from diversity transmission employing frequency or phase modulation, except that carriers would presumably require to be synchronized and that the advantage of frequency diversity would be lost.

Mr. G. Millington: The scattering law that leads to a high dependence on the frequency also implies a rapid decrease with increase of the scattering angle. This angle increases much more rapidly with lateral displacement from the mid-point of the path than with movement along the path. Experiments suggest that the falling-off of the signal with increasing scattering angle is not so severe as has been thought, and that it is advantageous to make the horizontal beam-width great enough to embrace the whole of the area in the scattering region visible from both the transmitter and receiver. This argument applies to distances near to the limiting distance, where the lateral angle increases rapidly as the distance decreases, but serious time delays will not occur from the sides of the illuminated region on the signal received from the mid-point.

Mr. Isted's experiments indicate that these considerations may affect the apparent gain of a given aerial array over the simple dipole and suggest a modification in the aerial design. I should like the authors' views on this possibility.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Mr. W. J. Bray, Dr. J. A. Saxton, Messrs. R. W. White and **G. W. Luscombe** (*in reply*): As Dr. Megaw points out, there has been considerable confusion concerning the application of the theory of turbulence to radio-wave propagation phenomena, and in particular concerning what exactly is meant by the scale of turbulence. We agree with his comments and are grateful for his clear exposition of the theory in question.

Dr. Smith-Rose draws attention to the possible serious effects of interference between communication circuits operating by means of ionospheric scattering and those operating in the more normal manner; there is no doubt that this is a very important problem which must be carefully considered by those responsible for the allocation of frequencies. He also stresses quite rightly the inefficient use of power in communication by scatter propagation; nevertheless, this may be the only means by which reliable radio communication can be maintained for nearly 100% of the time over paths of 1 000–2 000 km.

The siting of the receiving aerial in relation to the immediate foreground has a marked effect on the strength of the received signal, and the discrepancy in observed signal levels under different circumstances mentioned by Dr. Glazier is doubtless bound up with this fact. Dr. Glazier also raises the question of signal/noise ratio over the frequency band 30–60 Mc/s. No accurate evaluation of cosmic noise was made during the investigation, but there is little doubt that the overall signal/noise ratio, when all forms of noise are considered, is better at the lower frequencies.

Although the narrower bandwidth of 70 c/s suggested by Dr. Glazier may be adequate for frequency-shift telegraphy on high-frequency circuits, it might prove too narrow for scatter propagation where the random frequency shifts mentioned in Section 4.1 are present. Automatic error-correction equipment would probably enable lower aerial gains or transmitter power to be used, but the net improvement might not be as great as expected, since either a greater bandwidth for the same signalling speed, or a lower speed for the same bandwidth, would be required. The improvement to be gained by the use of frequency or aerial diversity is almost all obtained by their use at only one end of the circuit.

Both Mr. Hamer and Mr. Fitch comment on the apparent increase in the effective gain of the receiving aerial with increasing signal level shown in Fig. 11. If the data are replotted in the manner suggested by Mr. Fitch, with the signal strength observed with the dipole as a function of that observed with the array, and a line is drawn through the points based on the method of least squares, it is apparent that the signal level from the array increases more rapidly than that from the dipole: in fact, the effective gain of the array increases by about 1 dB for every increase of 3 dB in the signal strength as measured with the dipole. Mr. Fitch's more detailed analysis of the performance of directive aerials serves to underline the fact that some degradation in gain must generally occur, and also confirms the influence of the thickness of the scattering layer in relation to the region of overlap of the transmitting and receiving aerial beams.

In reply to Mr. Millington, it would appear that, for paths near to the limiting distance, the aerials should be designed in such a manner that the maximum scattering volume in the ionosphere is obtained. At shorter distances the beam widths would be determined by the greatest path differences, corresponding to the extremes of the scattering region, which can be tolerated.

We believe that the general characteristics of the observed signal show several significant correlations with meteoric activity. The number of bursts occurring is so large that, without a detailed statistical analysis, it is not possible to say whether the apparently systematic spacing of the bursts observed by Mr. Isted has any particular significance.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING), MARCH 1956

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.